



# Improved Approximation Algorithms for Bounded-Degree Local Hamiltonians

## Citation

Anshu, Anurag, David Gosset, Karen J. Morenz Korol, Mehdi Soleimanifar. "Improved Approximation Algorithms for Bounded-Degree Local Hamiltonians." *Phys. Rev. Lett.* 127, no. 25 (2021). DOI: 10.1103/physrevlett.127.250502

## Published version

<https://doi.org/10.1103/physrevlett.127.250502>

## Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37377980>

## Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles (OAP), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

## Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

## Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)

# Improved approximation algorithms for bounded-degree local Hamiltonians

Anurag Anshu<sup>1</sup>, David Gosset<sup>2</sup>, Karen J. Morenz Korol<sup>3</sup>, and Mehdi Soleimanifar<sup>4</sup>

<sup>1</sup> *Department of EECS & Challenge Institute for Quantum Computation, University of California, Berkeley, USA and Simons Institute for the Theory of Computing, Berkeley, California, USA.*

<sup>2</sup> *Department of Combinatorics and Optimization and Institute for Quantum Computing, University of Waterloo, Canada*

<sup>3</sup> *Department of Chemistry, University of Toronto, Canada and*

<sup>4</sup> *Center for Theoretical Physics, Massachusetts Institute of Technology, USA*

We consider the task of approximating the ground state energy of two-local quantum Hamiltonians on bounded-degree graphs. Most existing algorithms optimize the energy over the set of product states. Here we describe a family of shallow quantum circuits that can be used to improve the approximation ratio achieved by a given product state. The algorithm takes as input an  $n$ -qubit product state  $|v\rangle$  with mean energy  $e_0 = \langle v|H|v\rangle$  and variance  $\text{Var} = \langle v|(H - e_0)^2|v\rangle$ , and outputs a state with an energy that is lower than  $e_0$  by an amount proportional to  $\text{Var}^2/n$ . In a typical case, we have  $\text{Var} = \Omega(n)$  and the energy improvement is proportional to the number of edges in the graph. When applied to an initial random product state, we recover and generalize the performance guarantees of known algorithms for bounded-occurrence classical constraint satisfaction problems. We extend our results to  $k$ -local Hamiltonians and entangled initial states.

Quantum computers are capable of efficiently computing the dynamics of quantum many-body systems [1], and it is anticipated that they can be useful for scientific applications in physics, materials science and quantum chemistry. The extent of the quantum advantage for other important simulation tasks, such as computing low temperature properties of quantum systems, is still unknown. In this paper we consider the task of approximating the ground state energy of local Hamiltonians. Here it is natural to expect some improvement over classical machines which cannot even store the state of such systems efficiently. Indeed, classical methods such as the mean-field or Hartree-Fock approximations do not capture the entanglement structure present in the true ground state.

Motivated by small quantum computers that may be available in the near future, there has been increased interest in devising algorithms that consume few quantum resources and can be implemented across a wide range of hardware platforms. In this vein, heuristic algorithms for ground state preparation have been proposed based on variationally minimizing the energy over the output states of shallow (low-depth) quantum circuits [2–4]. Although variational algorithms have been rigorously analyzed for specific problems and some limitations are known [5–8], no general treatment of their efficacy exists. Characterizing the advantage offered by shallow quantum circuits and variational quantum algorithms stands as a pressing challenge.

In this paper, we derive rigorous bounds on the performance of shallow quantum circuits in estimating the ground state energy of local Hamiltonians. For simplicity, we state our results for a system of qubits with two-local interactions. In the Supplemental Material, we discuss extensions of our results to  $k$ -local Hamiltonians.

To begin, let  $G = (V, E)$  be a graph, and consider a

Hamiltonian

$$H = \sum_{\{i,j\} \in E} h_{ij} \quad (1)$$

with  $n = |V|$  qubits and nearest-neighbor interactions  $h_{ij}$  that act nontrivially only on qubits  $\{i, j\}$  at vertices connected by an edge. We assume without loss of generality that  $\|h_{ij}\| \leq 1$ . We are interested in the problem of approximating the ground energy or smallest eigenvalue  $\lambda_{\min}(H)$  of the Hamiltonian. It will be convenient to instead approximate the largest eigenvalue  $\lambda_{\max}(H)$ ; this convention matches the one used in classical optimization and is without loss of generality, since  $\lambda_{\min}(H) = -\lambda_{\max}(-H)$ . In the worst case, the problem of estimating the largest eigenvalue  $\lambda_{\max}(H)$  of Eq. (1) to within an additive error scaling inverse polynomially with  $n$  is believed to be intractable for quantum or classical computers [9]. Here we consider the approximation task where the goal is to compute an estimate  $e \leq \lambda_{\max}(H)$  such that the approximation ratio  $r \equiv e/\lambda_{\max}(H)$  is as large as possible. We will also be interested in efficient quantum algorithms that prepare states  $|\psi\rangle$  with good approximation ratios.

Besides describing local interactions encountered in physics, Hamiltonians of the form Eq. (1) can encode notable cost-functions considered in computer science and thus provide a physically motivated extension of the classical *approximation algorithm* setting [10]. For example, one may consider an Ising Hamiltonian for which  $h_{ij} = (I + Z_i Z_j)/2$ , where  $Z$  is the Pauli operator. This Hamiltonian is classical—that is, diagonal in the computational basis—and computing its maximum eigenvalue is equivalent to finding the Max-Cut of the graph  $G$ , a well-studied classical optimization problem. More generally, two-local quantum Hamiltonians may involve non-commuting terms such as Heisenberg interactions  $h_{ij} = 1/4(I - X_i X_j - Y_i Y_j - Z_i Z_j)$  (with Pauli  $X, Y$  and  $Z$  oper-

ators); the resulting optimization problem can be viewed as a quantum analogue of Max-Cut [11]. Quantum approximation algorithms aim to estimate the largest eigenvalue of such Hamiltonians and have been studied in several previous works. This includes the Heisenberg interactions mentioned above [11, 12] and more general settings in which the interaction terms  $h_{ij}$  are restricted to be positive semidefinite [13–15], or traceless [16, 17].

Despite considerable interest, the ultimate limits of efficient algorithms for quantum approximation algorithms are poorly understood. Approximation ratios approaching 1 are only known to be achievable for certain special families of graphs, including lattices or bounded-degree planar graphs using tensor product of  $O(1)$ -qubit states [18] or high degree graphs using tensor products of single-qubit states [13, 18, 19]. In certain cases, one may ascertain limitations on efficient achievable approximation ratios from the classical Probabilistically Checkable Proof (PCP) theorem [20–22], though stronger and more general limitations may follow from the quantum PCP conjecture if some version of it can be proven [23].

A quantum approximation algorithm typically outputs an estimate of the form  $\langle v|H|v\rangle$  where  $|v\rangle$  is a quantum state computed by the algorithm. A central challenge is to understand the structure of quantum states  $|v\rangle$  that achieve high approximation ratios in the general case. Most existing algorithms are based on tensor products of one- or few-qubit states, while Ref. [12] also considers states prepared by shallow quantum circuits. In this work we describe conditions under which the performance of such algorithms can be *improved*. We restrict our attention to local Hamiltonians on bounded-degree graphs and consider an improvement strategy based on shallow quantum circuits.

*Improvement of product states* To this end, suppose we are given an  $n$ -qubit state  $|v\rangle$  and a Hamiltonian Eq. (1) defined on a graph  $G = (V, E)$  with maximum degree  $d \geq 2$ . It will be convenient to assume (without loss of generality) that  $G$  is  $d$ -regular—we can ensure this by possibly adding some local terms  $h_{ij}$  which are equal to zero. We imagine that  $|v\rangle$  may be the output of some approximation algorithm such as the ones described above. Our aim is to efficiently compute a state with energy larger than  $\langle v|H|v\rangle$ . Moreover, we would like to increase this energy by an amount proportional to  $|E|$  in order to guarantee that the *approximation ratio* is larger by some additive constant. We show that this is possible if the following two conditions hold:

- (i) The variance of the energy, defined by

$$\text{Var}_v(H) = \langle v|H^2|v\rangle - \langle v|H|v\rangle^2,$$

satisfies  $\text{Var}_v(H) = \Omega(|E|)$  [24].

- (ii) The state  $|v\rangle$  is a product state. That is,  $|v\rangle =$

$|v_1\rangle \otimes |v_2\rangle \otimes \dots \otimes |v_n\rangle$  where each  $|v_i\rangle$  is a single-qubit state.

**Theorem 1.** *Given a product state  $|v\rangle$ , we can efficiently compute a depth- $(d+1)$  quantum circuit  $U$  such that the state  $|\psi\rangle = U|v\rangle$  satisfies*

$$\langle \psi|H|\psi\rangle \geq \langle v|H|v\rangle + \Omega\left(\frac{\text{Var}_v(H)^2}{d^2|E|}\right). \quad (2)$$

This result applies broadly to quantum optimization problems, but does not provide any improvement when specialized to the classical setting. To see this, note that condition (i) is not satisfied in the purely classical case where  $|v\rangle$  is a computational basis state and  $H$  is diagonal in the computational basis. Indeed, we have  $\text{Var}_v(H) = 0$  whenever  $|v\rangle$  is an eigenstate of  $H$ . On the other hand, condition (i) is fairly mild in the quantum setting, which can be seen from the following expression for the variance:

$$\text{Var}_v(H) = \sum_{\{i,j\} \cap \{k,l\} \neq \emptyset} (\langle v|h_{ij}h_{kl}|v\rangle - \langle v|h_{ij}|v\rangle \cdot \langle v|h_{kl}|v\rangle).$$

Since  $G$  is  $d$ -regular, the number of terms in the sum is  $O(d|E|)$ . So condition (i) is satisfied if the sum is proportional to the number of terms appearing in it.

Simple examples demonstrate that neither of the two conditions alone is enough to even guarantee the existence of a state with approximation ratio better than  $|v\rangle$  for large regular graphs. Condition (ii) alone is not sufficient because it is possible for a product state to have maximal energy  $\lambda_{\max}(H)$  (i.e., this occurs for all classical Hamiltonians). To see that condition (i) is not sufficient, one can consider the Max-Cut Hamiltonian on (say) an even cycle graph, and let  $|v\rangle$  be an equal superposition of two eigenstates of  $H$ , one with maximal energy  $|E|$  and one with energy  $|E| - \Theta(\sqrt{|E|})$ . The resulting state has approximation ratio  $1 - O(|E|^{-1/2})$  and variance  $\text{Var}_v(H) = \Omega(|E|)$ . Thus condition (i) is satisfied, but the approximation ratio cannot be improved by an additive constant.

In the special case where  $|v\rangle$  achieves the largest energy of any product state, we are able to strengthen the bound Eq. (2). We say that the product state  $|v\rangle$  is *locally optimal* for  $H$  if for any single-qubit Pauli  $Q$ , we have

$$\frac{d}{d\phi} \langle v|e^{-i\phi Q} H e^{i\phi Q}|v\rangle \Big|_{\phi=0} = 0,$$

or equivalently  $\langle v|[Q, H]|v\rangle = 0$ . As we show in the Supplemental Material, the bound in Eq. (2) can be improved to  $\langle v|H|v\rangle + \Omega(\frac{\text{Var}_v(H)^2}{d|E|})$  for locally optimal states.

Generally, however, the improvement stated in Eq. (2) is optimal in the sense that there exists a Hamiltonian

$H$  and a product state  $|v\rangle$  with  $\text{Var}_v(H) = \Theta(|E|)$  for which

$$\lambda_{\max}(H) - \langle v|H|v\rangle \leq O\left(\frac{\text{Var}_v(H)^2}{d^2|E|}\right). \quad (3)$$

For example, Eq. (3) is satisfied by the Hamiltonian with  $h_{ij} = Z_i + Z_j$  on any  $d$ -regular graph and the product state  $|v\rangle = (\cos(\theta)|0\rangle + \sin(\theta)|1\rangle)^{\otimes n}$ , for any  $\theta \in (0, \pi/2)$ . In this simple case, the right-hand side can be computed exactly and is equal to  $\frac{\text{Var}(H)^2}{d^2|E|} \cdot \frac{4\sin^2(\theta)}{\sin^2(2\theta)}$ .

To establish Theorem 1, we consider a variational family of states obtained from  $|v\rangle = \otimes_{i \in V} |v_i\rangle$  by applying a quantum circuit composed of nearest neighbor commuting gates on the interaction graph  $G$ . In particular, let  $P_1, P_2, \dots, P_n$  be any collection of single-qubit operators such that  $\|P_i\| \leq 1$  and

$$\langle v_i|P_i|v_i\rangle = 0 \quad \text{for all } i \in V.$$

Following [12], we define the circuit

$$V(\vec{\theta}) = \bigotimes_{\{i,j\} \in E} e^{i\theta_{ij}P_iP_j} = e^{i\sum_{\{i,j\} \in E} \theta_{ij}P_iP_j}. \quad (4)$$

Here,  $\vec{\theta}$  is an array of real parameters  $\{\theta_{ij}\}_{\{i,j\} \in E}$ . Since by assumption, the interaction graph  $G$  is  $d$ -regular, the quantum circuit  $V(\vec{\theta})$  can be implemented with circuit depth  $d+1$ . It is not hard to see that this variational family includes as a special case the level-1 Quantum Approximate Optimization Algorithm (QAOA) for 2-local classical Hamiltonians [3]. For a given choice of operators  $\{P_i\}_{i \in V}$ , the following theorem lower bounds the improvement in the energy after applying the the quantum circuit  $V(\vec{\theta})$  to  $|v\rangle$ .

**Theorem 2.** *Let  $|v\rangle$  be a product state and  $|\psi\rangle = V(\vec{\theta})|v\rangle$  be the state prepared by the quantum circuit Eq. (4). Define the positive real parameter  $\alpha$  by*

$$\alpha = \mathbb{E}_{\{i,j\} \in E} |\langle v_i, v_j | [P_i P_j, h_{ij}] | v_i, v_j \rangle|, \quad (5)$$

where the expectation is with respect to the uniform distribution over the edges. There is an efficient classical algorithm to select parameters  $\vec{\theta}$  satisfying

$$\langle \psi | H | \psi \rangle \geq \langle v | H | v \rangle + \Omega(|E|\alpha^2/d). \quad (6)$$

*Proof.* Write  $N_{ij}$  for the set of edges  $\{k, \ell\} \in E$  incident to a given edge  $\{i, j\} \in E$ . The latter edge is included as well, i.e.,  $\{i, j\} \in N_{ij}$ . Consider the energy of a term

$$\langle \psi | h_{ij} | \psi \rangle = \langle v | V(\vec{\theta})^\dagger h_{ij} V(\vec{\theta}) | v \rangle.$$

The gates in  $V(\vec{\theta})$  which are associated with edges that are not incident with  $\{i, j\}$  can be cancelled, leaving

$\langle v | V_{ij}^\dagger h_{ij} V_{ij} | v \rangle$  where  $V_{ij} = \prod_{\{k, \ell\} \in N_{ij}} e^{i\theta_{k\ell} P_k P_\ell}$ . Thus

$$\begin{aligned} \langle \psi | h_{ij} | \psi \rangle &= \langle v | h_{ij} | v \rangle + \\ &\sum_{m=2}^{\infty} \frac{i^m}{m!} \langle v | \left[ \sum_{\{k, \ell\} \in N_{ij}} -\theta_{k\ell} P_k P_\ell, h_{ij} \right]_m | v \rangle. \end{aligned} \quad (7)$$

Here,  $[A, B]_m$  is the  $m$ -nested commutator  $[A, [A, \dots [A, B] \dots]]$ . Using the fact that  $\langle v_k | P_k | v_k \rangle = 0$  for all  $k$ , the  $m=1$  term simplifies to

$$\sum_{\{k, \ell\} \in N_{ij}} -i\theta_{k\ell} \langle v | [P_k P_\ell, h_{ij}] | v \rangle = -i\theta_{ij} \langle v | [P_i P_j, h_{ij}] | v \rangle. \quad (8)$$

At this stage, we make the choice

$$\theta_{ij} = \theta \cdot \text{sign}(-i \langle v | [P_i P_j, h_{ij}] | v \rangle), \quad (9)$$

where the parameter  $\theta$  will be determined later. Substituting in Eq. (8) gives

$$\sum_{\{k, \ell\} \in N_{ij}} -i\theta_{k\ell} \langle v | [P_k P_\ell, h_{ij}] | v \rangle = \theta |\langle v_i, v_j | [P_i P_j, h_{ij}] | v_i, v_j \rangle|. \quad (10)$$

For  $m > 1$ , we have

$$\begin{aligned} &\left| \langle v | \left[ \sum_{\{k, \ell\} \in N_{ij}} -\theta_{k\ell} P_k P_\ell, h_{ij} \right]_m | v \rangle \right| \\ &\leq \sum_{\substack{\{k_1, \ell_1\}, \{k_2, \ell_2\}, \dots \\ \{k_m, \ell_m\} \in N_{ij}}} \theta^m |\langle v | [P_{k_1} P_{\ell_1}, [\dots, [P_{k_m} P_{\ell_m}, h_{ij}]]] | v \rangle|. \end{aligned}$$

The only nonzero terms are those in which the expression  $\langle v_s | P_s | v_s \rangle$  does not appear. To upper bound the number of nonzero terms, we count the number of tuples  $(\{k_1, \ell_1\}, \{k_2, \ell_2\}, \dots, \{k_m, \ell_m\})$  such that no vertex in  $V \setminus \{i, j\}$  appears exactly once. An upper bound is provided in the following.

**Claim 1.** *Let  $m \geq 2$ . The number of ordered tuples of edges  $(\{k_1, \ell_1\}, \{k_2, \ell_2\}, \dots, \{k_m, \ell_m\}) \in N_{ij}^{\times m}$  in which no vertex in  $V \setminus \{i, j\}$  appears exactly once is at most  $(2m\sqrt{d})^m$ .*

*Proof.* First, we count all such tuples  $(\{k_1, \ell_1\}, \{k_2, \ell_2\}, \dots, \{k_m, \ell_m\})$  in which the edge  $\{i, j\}$  does not appear. Each one can be generated by choosing a tuple of vertices  $(v_1, v_2, \dots, v_m)$  incident to  $\{i, j\}$  and then specifying a neighbor, either  $i$  or  $j$ , for each of them. An upper bound is obtained by counting the number of tuples  $(v_1, v_2, \dots, v_m)$  such that each  $v_p$  occurs at least twice and then multiplying by  $2^m$ . Any tuple  $(v_1, v_2, \dots, v_m)$  of this form can be generated as follows. First, for each  $i = 1, 2, \dots, m$  we choose a color  $c(i) \in \{1, 2, \dots, m/2\}$ . We set  $v_k = v_{k'}$  whenever  $c(k) = c(k')$ . We then assign a neighbor of  $i$  or  $j$  to each color  $\{1, 2, \dots, m/2\}$ . Since vertices  $i$  and  $j$  each

have at most  $d$  neighbors, we see that the number of tuples  $(v_1, v_2, \dots, v_m)$  such that each  $v_p$  occurs at least twice is at most  $(m/2)^m \cdot (2d)^{m/2}$ . The number of tuples of edges  $(\{k_1, \ell_1\}, \{k_2, \ell_2\}, \dots, \{k_m, \ell_m\}) \in N_{ij}^{\times m}$  in which no vertex in  $V \setminus \{i, j\}$  appears exactly once, and the edge  $\{i, j\}$  does not occur, is then at most  $2^m \cdot (m/2)^m \cdot (2d)^{m/2}$ .

In order to account for the appearance of the edge  $\{i, j\}$ , we fix the number of places  $u$  where the edge appears and then count as before for the  $m-u$  places. This number is

$$\sum_{u=0}^m \binom{m}{u} ((m-u)\sqrt{2d})^{m-u} \leq (2m\sqrt{d})^m.$$

□

Finally, using Eq. (9) and the fact that  $\|h_{ij}\|, \|P_i\| \leq 1$ , we can upper bound

$$\theta^m |\langle v | [P_{k_1} P_{\ell_1}, [\dots, [P_{k_m} P_{\ell_m}, h_{ij}]]] | v \rangle| \leq (2\theta)^m.$$

Thus, the sum of all  $m > 1$  terms in Eq. (7) has magnitude at most

$$\sum_{m=2}^{\infty} \frac{1}{m!} (4m\sqrt{d})^m \theta^m \leq \sum_{m=0}^{\infty} (4e\sqrt{d}\theta)^{m+2} \leq 32e^2 d \theta^2$$

assuming  $\theta \leq \frac{1}{8e\sqrt{d}}$  (where we used the bound  $m^m/m! \leq e^m$ ). Combining with Eqs. (7,10) and summing over all  $\{i, j\} \in E$ , we get

$$\langle \psi | H | \psi \rangle \geq \langle v | H | v \rangle + |E| (\theta\alpha - 32e^2 d \theta^2).$$

We may then choose  $\theta = O(\alpha/d)$  to get the desired lower bound. □

Let us now see how Theorem 1 is obtained as a consequence of Theorem 2. The lower bound (6) applies to any choice of operators  $\{P_i\}_{i \in V}$ . We will choose these operators in a way that gives the variance bound Eq. (2). In the following, for convenience and without loss of generality, we shall work in a local basis in which our initial product state is  $|v\rangle \equiv |0^n\rangle$ . Our starting point is the observation that the variance of a 2-local Hamiltonian can be expressed in this basis as

$$\text{Var}_v(H) = \langle 0^n | H Q_1 H | 0^n \rangle + \langle 0^n | H Q_2 H | 0^n \rangle,$$

where  $Q_t$  is the projector onto computational basis states with Hamming weight  $t \in \{1, 2\}$ . This implies that

$$\langle 0^n | H Q_t H | 0^n \rangle \geq \text{Var}_v(H)/2 \quad (11)$$

for some  $t \in \{1, 2\}$ . Suppose  $t = 2$  and let  $X_i, Y_i$ , and  $Z_i$  be the Pauli operators. We define  $\alpha_1$  to be the RHS of Eq. (5) with  $P_i = X_i$  for all  $i$ , and similarly  $\alpha_2$  with

$P_i = (X_i + Y_i)/\sqrt{2}$  for all  $i$ . By a direct calculation we see that

$$\begin{aligned} \alpha_1 &= \frac{2}{|E|} \sum_{\{i,j\} \in E} |\text{Im}(\langle 11 | h_{ij} | 00 \rangle)| \\ \alpha_2 &= \frac{2}{|E|} \sum_{\{i,j\} \in E} |\text{Re}(\langle 11 | h_{ij} | 00 \rangle)| \end{aligned} \quad (12)$$

and therefore

$$\langle 0^n | H Q_2 H | 0^n \rangle = \sum_{\{i,j\} \in E} |\langle 11 | h_{ij} | 00 \rangle|^2 \leq |E| \left( \frac{\alpha_1 + \alpha_2}{2} \right).$$

This means  $\max\{\alpha_1, \alpha_2\} \geq |E|^{-1} \langle 0^n | H Q_2 H | 0^n \rangle$  which together with Eq. (11) implies that when  $t = 2$ , we can efficiently find a series of operators  $P_i$  such that the parameter  $\alpha$  satisfies  $\alpha \geq (2|E|)^{-1} \text{Var}_v(H)$ . By plugging this in Eq. (6), we obtain  $\langle \psi | H | \psi \rangle \geq \langle v | H | v \rangle + \Omega(\frac{\text{Var}_v(H)}{d|E|})$ . Thus if  $t = 2$  we obtain a better lower bound than the one claimed in Theorem 1. Otherwise, if  $t = 1$ , then a simple calculation (reproduced in the Supplemental Material) shows that one can efficiently compute a product state with energy at least  $\langle v | H | v \rangle + \Omega(\frac{\text{Var}(H)}{d^2|E|})$ . In general, the choice between  $t = 1$  and  $t = 2$  can be efficiently determined. Thus we obtain Theorem 1. In the Supplemental Material, we show that if  $|v\rangle$  is locally optimal for  $H$ , then  $\langle 0^n | H Q_1 H | 0^n \rangle = 0$  and  $t = 2$ , so we obtain the better bound described above.

Let us briefly illustrate how these results can be applied to the quantum Max-Cut Hamiltonian considered in Refs. [11, 12]. The Hamiltonian is built from local terms  $h_{ij} = w_{ij} \Pi_{ij}$ , where  $0 \leq w_{ij} \leq 1$  and  $\Pi_{ij} = (I - X_i X_j - Y_i Y_j - Z_i Z_j)/4$  is the projector onto the antisymmetric state of two qubits. This Hamiltonian has the special feature that any product state  $|v\rangle$  is locally optimal, and moreover, we have  $|\langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle| = \langle v_i, v_j | h_{ij} | v_i, v_j \rangle$ . Therefore

$$\text{Var}_v(H) = \sum_{\{i,j\} \in E} \langle v | h_{ij} | v \rangle^2 \geq |E|^{-1} \langle v | H | v \rangle^2$$

using Cauchy-Schwarz. We may then efficiently compute a state  $|\psi\rangle$  such that

$$\langle \psi | H | \psi \rangle \geq \langle v | H | v \rangle + \Omega\left(\frac{\langle v | H | v \rangle^4}{d|E|^3}\right). \quad (13)$$

We see that if the initial state has approximation ratio  $\langle v | H | v \rangle / |E| = r$  then the state  $|\psi\rangle$  improves this to  $r + \Omega(r^4/d)$  [25].

The preceding example demonstrates the power of Theorem 1 and shows that for the quantum Max-Cut problem, the approximation ratio of *any* product state can be improved by applying a shallow quantum circuit. For more general two-local Hamiltonians, we can guarantee an improvement in the approximation ratio whenever

the condition  $\text{Var}_v(H) = \Omega(|E|)$  holds, which we expect for most (but not all) product states. Below we discuss two natural extensions of our results. First, we ask whether one can improve approximation ratios attained by more general families of quantum states. Along these lines, we provide an extension of Theorem 1 to the more general case where  $|v\rangle$  is any state prepared by a quantum circuit of depth  $D = O(1)$ . Next, we show how one can improve the approximation ratio achieved by a random product state  $|v\rangle$ . Using Theorem 2, we show that the approximation ratio can be improved by  $\Omega(1/d)$  for any Hamiltonian with nontrivial two-local interactions, and by  $\Omega(1/\sqrt{d})$  if the interaction graph is triangle-free.

*Improvement of bounded-depth states* Recall that for any  $n$ -qubit quantum circuit and any qubit  $j \in [n]$ , we may define the lightcone  $\mathcal{L}(j) \subseteq [n]$  which consists of all output qubits that are causally connected to  $j$ . Define the maximum lightcone size  $\ell = \max_{j \in [n]} |\mathcal{L}(j)|$ . We have  $\ell \leq 2^D$  for any depth  $D$  circuit composed of two-qubit gates.

**Theorem 3.** *Let  $|v\rangle = W|0^n\rangle$  where  $W$  is a quantum circuit with maximum lightcone size  $\ell$ . There is an efficient classical algorithm that computes a quantum circuit  $U$  such that  $|\psi\rangle = U|v\rangle$  satisfies*

$$\langle \psi | H | \psi \rangle = \langle v | H | v \rangle + \Omega\left(\frac{\text{Var}(H)^2}{\ell^{10} d^2 |E|}\right).$$

For constant depth circuits we have  $\ell = O(1)$  and we get the same asymptotic energy improvement as we established previously in Theorem 1 for product states. However, in this case the circuit  $U$  that we construct is not constant-depth. In the Supplemental Material, we show that the improvement stated above can also be obtained for states  $|v\rangle$  that are the unique ground states of a gapped local Hamiltonian  $F$ . In that case,  $\ell$  is replaced by the locality of the Hamiltonian  $F$ . Thus, Theorem 3 extends to a broad class of tensor network states (such as PEPS of low bond dimension) that have a gapped parent Hamiltonian.

The theorem provides limitations on the energy that can be achieved by any state  $|v\rangle$  produced by a bounded-depth circuit. In particular, since  $\langle \psi | H | \psi \rangle \leq \lambda_{\max}(H)$ , we find that

$$\langle v | H | v \rangle \leq \lambda_{\max}(H) - \Omega\left(\frac{\text{Var}(H)^2}{\ell^{10} d^2 |E|}\right).$$

This shows that the approximation ratio achievable by constant-depth states  $|v\rangle$  with  $\text{Var}(H) = \Omega(|E|)$  is bounded away from 1. An interesting direction for future work is to explore whether one can use this fact to exhibit new local Hamiltonian systems with the almost-linear NLTS (No Low-energy Trivial States) property [26, 27].

*Improvement of random assignments* Given an instance of a (classical) constraint satisfaction problem,

one may consider the trivial algorithm in which each variable is chosen independently and uniformly at random. Remarkably, efficient algorithms which improve over the approximation ratio achieved by this simple strategy are not likely to exist in the general case [28]. On the other hand, for structured cases such as bounded-degree graphs, improvement is possible. In particular, on degree- $d$  graphs, one can efficiently find an assignment satisfying a  $\mu + \Omega(\frac{1}{d})$  fraction of constraints [29]. Here  $\mu$  is the expected fraction of constraints satisfied by a uniformly random assignment. It has been shown that when a degree- $d$  graph is triangle-free, there are efficient “local” algorithms that find a binary string satisfying a  $\mu + \Omega(\frac{1}{\sqrt{d}})$  fraction of constraints by starting with a uniformly random assignment [30, 31] or quantum superposition [32] and then locally updating each bit/qubit as a function of the state of its neighbors.

Below we show that this optimal dependence on  $d$  can be recovered and generalized to the local Hamiltonian setting by applying our algorithm in Theorem 2 to a randomly chosen product state. For randomly chosen  $|v\rangle$ , the parameter  $\alpha$  in Theorem 2 can be related to the 2-norm of the quadratic terms in the Pauli expansion of the Hamiltonian. More precisely, for an  $n$ -qubit operator  $O = \sum_{i < j} \sum_{x,y} f_{xy}^{ij} \sigma_x^i \otimes \sigma_y^j$  where  $\sigma_0 = I$  and  $\{\sigma_1, \sigma_2, \sigma_3\}$  are the Pauli matrices, we define

$$\text{quad}(O) = \sum_{i < j} \sum_{x > 0, y > 0} (f_{xy}^{ij})^2.$$

**Theorem 4.** *There is an efficient randomized algorithm which computes a depth- $d + 1$  circuit  $U$  such that  $|\psi\rangle = U|v\rangle$  satisfies*

$$\mathbb{E}_v \langle \psi | H | \psi \rangle \geq \mathbb{E}_v \langle v | H | v \rangle + \Omega\left(\frac{\text{quad}(H)^2}{d|E|}\right).$$

*If the graph is triangle-free then the right-hand side can be replaced with  $\mathbb{E}_v \langle v | H | v \rangle + \Omega\left(\frac{\text{quad}(H)}{\sqrt{d}}\right)$ .*

The proof of Theorem 4 is provided in the Supplemental Material. We also show that for triangle-free graphs one can efficiently compute *product* states matching the approximation ratios quoted above using a local classical algorithm similar to the ones described in Refs. [30, 31]. Thus, low depth quantum circuits are not necessary to achieve the asymptotic  $\Omega(1/\sqrt{d})$  scaling in this case. Nevertheless, one may take the output product state of such algorithms and improve it further using the shallow quantum circuit from Theorem 1.

*Discussion* For local Hamiltonian problems on bounded-degree graphs, we showed that the approximation ratio achieved by a product state can be improved by a shallow quantum circuit, assuming a mild condition on its variance. Our quantum algorithm generalizes the level-1 Quantum Approximate Optimization Algorithm

(QAOA) and extends its applicability beyond classical cost functions. By applying our algorithm to randomly chosen product states we generalized known algorithms for bounded-occurrence classical constraint satisfaction problems. Our results quantify the improvement that shallow quantum circuits can provide over methods based on product states.

*Acknowledgments* AA acknowledges support from the NSF QLCI program through grant number OMA-2016245. DG acknowledges the support of the Natural Sciences and Engineering Research Council of Canada through grant number RGPIN-2019-04198, the Canadian Institute for Advanced Research, and IBM Research. KJMK acknowledges support from NSERC Vanier Canada Graduate Scholarship. MS was supported by NSF grant CCF-1729369, a Samsung Advanced Institute of Technology Global Research Cluster and grant number FXQi-RFP-1811A from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation.

- 
- [1] Seth Lloyd, “Universal quantum simulators,” *Science*, 1073–1078 (1996).
- [2] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O’Brien, “A variational eigenvalue solver on a photonic quantum processor,” *Nature communications* **5**, 4213 (2014).
- [3] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, “A quantum approximate optimization algorithm,” *arXiv preprint arXiv:1411.4028* (2014).
- [4] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” *Nature* **549**, 242–246 (2017).
- [5] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature communications* **9**, 1–6 (2018).
- [6] Edward Farhi, David Gamarnik, and Sam Gutmann, “The quantum approximate optimization algorithm needs to see the whole graph: A typical case,” (2020), [arXiv:2004.09002](https://arxiv.org/abs/2004.09002) [quant-ph].
- [7] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang, “Obstacles to variational quantum optimization from symmetry protection,” *Phys. Rev. Lett.* **125**, 260505 (2020).
- [8] Sergey Bravyi, David Gosset, and Ramis Movassagh, “Classical algorithms for quantum mean values,” *Nature Physics* **17**, 337–341 (2021).
- [9] In particular, a decision version of this problem is complete for the complexity class QMA which is a quantum generalization of NP.
- [10] Vijay V Vazirani, *Approximation algorithms* (Springer Science & Business Media, 2013).
- [11] Sevag Gharibian and Ojas Parekh, “Almost optimal classical approximation algorithms for a quantum generalization of Max-Cut,” in *Approximation, randomization, and combinatorial optimization. Algorithms and techniques, LIPIcs. Leibniz Int. Proc. Inform., Vol. 145* (Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019) pp. Art. No. 31, 17.
- [12] Anurag Anshu, David Gosset, and Karen Morenz, “Beyond product state approximations for a quantum analogue of max cut,” in *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)* (Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020).
- [13] Sevag Gharibian and Julia Kempe, “Approximation algorithms for qma-complete problems,” *SIAM Journal on Computing* **41**, 1028–1050 (2012).
- [14] Sean Hallgren, Eunou Lee, and Ojas Parekh, “An approximation algorithm for the max-2-local hamiltonian problem,” in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)* (Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020).
- [15] Ojas Parekh and Kevin Thompson, “Beating random assignment for approximating quantum 2-local hamiltonian problems,” (2020), [arXiv:2012.12347](https://arxiv.org/abs/2012.12347) [quant-ph].
- [16] Aram W Harrow and Ashley Montanaro, “Extremal eigenvalues of local hamiltonians,” *Quantum* **1**, 6 (2017).
- [17] Sergey Bravyi, David Gosset, Robert König, and Kristan Temme, “Approximation algorithms for quantum many-body problems,” *Journal of Mathematical Physics* **60**, 032203 (2019).
- [18] Nikhil Bansal, Sergey Bravyi, and Barbara M Terhal, “Classical approximation schemes for the ground-state energy of quantum and classical ising spin hamiltonians on planar graphs,” *arXiv preprint arXiv:0705.1115* (2007).
- [19] Fernando GSL Brandao and Aram W Harrow, “Product-state approximations to quantum states,” *Communications in Mathematical Physics* **342**, 47–80 (2016).
- [20] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy, “Proof verification and the hardness of approximation problems,” *J. ACM* **45**, 501–555 (1998).
- [21] Sanjeev Arora and Shmuel Safra, “Probabilistic checking of proofs: a new characterization of NP,” *J. ACM* **45**, 70–122 (1998).
- [22] Irit Dinur, “The PCP theorem by gap amplification,” *J. ACM* **54**, Art. 12, 44 (2007).
- [23] Dorit Aharonov, Itai Arad, and Thomas Vidick, “Guest column: the quantum pcp conjecture,” *Acm sigact news* **44**, 47–79 (2013).
- [24] I.e. there is a universal constant  $c$ , such that asymptotically  $\text{Var}_v(H) \geq c \cdot |E|$ .
- [25] A better bound can be obtained by directly computing the parameter  $\alpha$  for a randomized choice of operators  $\{P_i\}$ . In that case,  $\mathbb{E}\alpha \geq \Omega(\langle v|H|v\rangle/|E|)$  which results in an improvement of  $\Omega(\langle v|H|v\rangle^2/(d|E|))$ .
- [26] Anurag Anshu and Chinmay Nirkhe, “Circuit lower bounds for low-energy states of quantum code hamiltonians,” (2021), [arXiv 2011.02044](https://arxiv.org/abs/2011.02044).
- [27] Michael H. Freedman and Matthew B. Hastings, “Quantum systems on non-k-hyperfinite complexes: A generalization of classical statistical mechanics on expander graphs,” *Quantum Info. Comput.* **14**, 144–180 (2014).

- [28] Johan Haastad, “Some optimal inapproximability results,” *J. ACM* **48**, 798–859 (2001).
- [29] Johan Haastad, “On bounded occurrence constraint satisfaction,” *Inform. Process. Lett.* **74**, 1–6 (2000).
- [30] Boaz Barak, Ankur Moitra, Ryan O’Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright, “Beating the random assignment on constraint satisfaction problems of bounded degree,” (2015), [arXiv:1505.03424 \[cs.CC\]](https://arxiv.org/abs/1505.03424).
- [31] Matthew B Hastings, “Classical and quantum bounded depth approximation algorithms,” arXiv preprint [arXiv:1905.07047](https://arxiv.org/abs/1905.07047) (2019).
- [32] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, “A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem,” (2015), [arXiv:1412.6062 \[quant-ph\]](https://arxiv.org/abs/1412.6062).
- [33] Ryan O’Donnell, *Analysis of boolean functions* (Cambridge University Press, 2014).
-



## Supplemental Materials

### IMPROVEMENT OF PRODUCT STATES

In this section we provide the full details of the proof of Theorem 1. It will be convenient to work in a local basis defined by  $|v\rangle$ , such that  $|v\rangle = |0^n\rangle$  and

$$\text{Var}_v(H) = \langle 0^n | H^2 | 0^n \rangle - (\langle 0^n | H | 0^n \rangle)^2.$$

For ease of notation we write  $\text{Var}(H) = \text{Var}_v(H)$ . Recall the quantity  $\alpha$  defined in Eq. (5):

$$\alpha = \mathbb{E}_{\{i,j\} \in E} |\langle v_i, v_j | [P_i P_j, h_{ij}] | v_i, v_j \rangle|, \quad (\text{S1})$$

We will use the following proposition.

**Proposition 1.** *Let  $Q_2$  be the projector onto computational basis states with Hamming weight 2. We can efficiently choose operators  $\{P_i\}_{i \in V}$  such that*

$$\alpha \geq \frac{1}{|E|} \cdot \langle 0^n | H Q_2 H | 0^n \rangle. \quad (\text{S2})$$

*Proof.* Let  $\alpha_1$  be Eq. (5) with  $P_i = X_i$  for all  $i$ , and let  $\alpha_2$  be Eq. (5) with  $P_i = (X_i + Y_i)/\sqrt{2}$  for all  $i$ . Direct calculation shows that

$$\begin{aligned} \alpha_1 &= \frac{2}{|E|} \sum_{\{i,j\} \in E} |\text{Im}(\langle 11 |_{ij} \otimes \langle 0^{n-2} | h_{ij} | 0^n \rangle)| \\ \alpha_2 &= \frac{2}{|E|} \sum_{\{i,j\} \in E} |\text{Re}(\langle 11 |_{ij} \otimes \langle 0^{n-2} | h_{ij} | 0^n \rangle)| \end{aligned} \quad (\text{S3})$$

We can express  $\langle 0^n | H Q_2 H | 0^n \rangle$  as

$$\begin{aligned} \langle 0^n | H Q_2 H | 0^n \rangle &= \sum_{\{i,j\} \in E} |\langle 11 |_{ij} \otimes \langle 0^{n-2} | h_{ij} | 0^n \rangle|^2 \\ &= \sum_{\{i,j\} \in E} (\text{Im}(\langle 11 |_{ij} \otimes \langle 0^{n-2} | h_{ij} | 0^n \rangle))^2 + (\text{Re}(\langle 11 |_{ij} \otimes \langle 0^{n-2} | h_{ij} | 0^n \rangle))^2 \\ &\leq \sum_{\{i,j\} \in E} |\text{Im}(\langle 11 |_{ij} \otimes \langle 0^{n-2} | h_{ij} | 0^n \rangle)| + |\text{Re}(\langle 11 |_{ij} \otimes \langle 0^{n-2} | h_{ij} | 0^n \rangle)| \\ &= |E| \cdot \frac{\alpha_1 + \alpha_2}{2}, \end{aligned}$$

where we used the fact that  $\|h_{ij}\| \leq 1$  in going from the second to the third line above. Now the last line implies that either  $\alpha_1$  or  $\alpha_2$  achieves the bound from Eq. (S2). Moreover, the choice of  $\alpha_1$  or  $\alpha_2$  can be efficiently determined.  $\square$

*Proof of Theorem 1.* Let  $Q_t$  be the projector onto computational basis states with Hamming weight  $t \in \{1, 2\}$ . Since  $H$  is two-local we have

$$\text{Var}(H) = \langle 0^n | H Q_1 H | 0^n \rangle + \langle 0^n | H Q_2 H | 0^n \rangle.$$

Therefore  $\langle 0^n | H Q_t H | 0^n \rangle \geq \text{Var}(H)/2$  for some  $t \in \{1, 2\}$ . If  $t = 2$  then we may use Proposition 1 which gives

$$\max\{\alpha_1, \alpha_2\} \geq \frac{1}{2|E|} \text{Var}(H).$$

Combining this with Theorem 2, we arrive at

$$\langle \psi | H | \psi \rangle \geq \langle 0^n | H | 0^n \rangle + \Omega\left(\frac{\text{Var}(H)^2}{d|E|}\right)$$

which is better than the desired lower bound.

Next suppose  $\langle 0^n | H Q_1 H | 0^n \rangle \geq \text{Var}(H)/2$ . Define

$$L = \sum_{j=1}^n (-1)^{a_j} P_j$$

where each  $P_j$  is a single-qubit Pauli operator acting nontrivially only on qubit  $j$ , and  $a_j \in \{0, 1\}$  is chosen so that

$$i \langle 0^n | (-1)^{a_j} [P_j, H] | 0^n \rangle = |\langle 0^n | [P_j, H] | 0^n \rangle|.$$

Define  $|\theta\rangle = e^{-i\theta L} | 0^n \rangle$  where  $\theta$  is a real parameter that we will fix later. Then

$$\langle \theta | H | \theta \rangle = \langle 0^n | H | 0^n \rangle + \theta \sum_{j=1}^n |\langle 0^n | [P_j, H] | 0^n \rangle| + \text{Err},$$

where

$$\begin{aligned} |\text{Err}| &= \left| \sum_{m \geq 2} \frac{i^m \theta^m}{m!} \langle 0^n | [L, H]_m | 0^n \rangle \right| \\ &\leq |E| \sum_{m \geq 2} \frac{\theta^m 4^m}{m!} \\ &\leq 16\theta^2 |E| e^{4\theta}. \end{aligned}$$

In the second line we used the fact that

$$[L, h_{ij}]_m = [(-1)^{a_i} P_i + (-1)^{a_j} P_j, h_{ij}]_m$$

can be expanded as a sum of  $2^m$  terms each of norm at most  $2^m$ . Now define

$$\beta = \frac{1}{|E|} \sum_{j=1}^n |\langle 0^n | [P_j, H] | 0^n \rangle|. \quad (\text{S4})$$

and note that since  $\|[P_j, H]\| \leq 2d$  for all  $j$  we have

$$\beta \leq \frac{1}{|E|} \sum_{j=1}^n 2d \leq 4.$$

Then

$$\langle \theta | H | \theta \rangle \geq \langle 0^n | H | 0^n \rangle + |E| (\theta\beta - 16\theta^2 e^{4\theta}).$$

Choosing  $\theta = \beta/32$  gives

$$\begin{aligned} \langle \theta | H | \theta \rangle &\geq \langle 0^n | H | 0^n \rangle + |E| \left( \frac{\beta^2}{32} - \frac{\beta^2}{64} e^{\beta/8} \right) \\ &\geq \langle 0^n | H | 0^n \rangle + |E| \left( \frac{\beta^2}{32} - \frac{\beta^2}{64} e^{1/2} \right) \\ &\geq \langle 0^n | H | 0^n \rangle + 0.001 \cdot |E| \beta^2. \end{aligned} \quad (\text{S5})$$

Now let  $\beta_1$  be given by Eq. (S4) with  $P_i = X_i$  for all  $i$ , and let  $\beta_2$  be given by Eq. (S4) with  $P_i = Y_i$  for all  $i$ . Then

$$\begin{aligned}
\frac{\beta_1 + \beta_2}{2} &= \frac{1}{2|E|} \sum_{j=1}^n |\langle 0^n | [X_j, H] | 0^n \rangle| + |\langle 0^n | [Y_j, H] | 0^n \rangle| \\
&\geq \frac{1}{4d|E|} \sum_{j=1}^n |\langle 0^n | [X_j, H] | 0^n \rangle|^2 + |\langle 0^n | [Y_j, H] | 0^n \rangle|^2 \\
&= \frac{1}{4d|E|} \sum_{j=1}^n |2\text{Im}(\langle \hat{e}_j | H | 0^n \rangle)|^2 + |2\text{Re}(\langle \hat{e}_j | H | 0^n \rangle)|^2 \\
&= \frac{1}{d|E|} \langle 0^n | H Q_1 H | 0^n \rangle \\
&\geq \frac{1}{2d|E|} \text{Var}(H).
\end{aligned}$$

Therefore either  $\beta_1$  or  $\beta_2$  is larger than the RHS above. Plugging this into Eq. (S5) we arrive at

$$\langle \theta | H | \theta \rangle \geq \langle 0^n | H | 0^n \rangle + \Omega\left(\frac{\text{Var}(H)^2}{d^2|E|}\right).$$

□

Finally, let us discuss a special case in which the bound from Theorem 1 can be improved. We say that a product state  $|v\rangle$  is locally optimal for  $H$  if, for any single-qubit Pauli  $Q$  we have

$$\frac{d}{d\phi} \langle v | e^{-i\phi Q} H e^{i\phi Q} | v \rangle \Big|_{\phi=0} = 0,$$

or equivalently

$$\langle v | [Q, H] | v \rangle = 0. \tag{S6}$$

As in the above, for simplicity we shall work in a local basis defined by  $v$ , so that  $|v\rangle = |0^n\rangle$ .

**Claim 2.** *Suppose  $|0^n\rangle$  is locally optimal for  $H$ . Then for any string  $z \in \{0, 1\}^n$  with Hamming weight  $|z| = 1$  we have*

$$\langle z | H | 0^n \rangle = 0. \tag{S7}$$

*Proof.* Without loss of generality consider the case where  $z = 10^{n-1}$ . Then

$$|2\text{Im}(\langle z | H | 0^n \rangle)| = |\langle 0^n | [X_1, H] | 0^n \rangle| = 0 \quad \text{and} \quad |2\text{Re}(\langle z | H | 0^n \rangle)| = |\langle 0^n | [Y_1, H] | 0^n \rangle| = 0,$$

where we used Eq. (S6). □

**Claim 3.** *Suppose  $|0^n\rangle$  is locally optimal for  $H$ . We may efficiently choose  $\{P_i\}$  and  $\{\theta_{ij}\}$  such that*

$$\langle \psi | H | \psi \rangle \geq \langle 0^n | H | 0^n \rangle + \Omega\left(\frac{\text{Var}(H)^2}{d|E|}\right). \tag{S8}$$

*Proof.* Since  $H$  is two-local we have

$$\text{Var}(H) = \langle 0^n | H Q_1 H | 0^n \rangle + \langle 0^n | H Q_2 H | 0^n \rangle = \langle 0^n | H Q_2 H | 0^n \rangle,$$

where in the last equality we used claim 2. The claim then follows directly by combining Proposition 1 and Theorem 2. □

## IMPROVEMENT OF RANDOM STATES

We prove the first part of Theorem 4 regarding general degree- $d$  graphs, which is implied by the following lemma.

**Lemma 1.** *Let  $|v\rangle = |v_1\rangle \otimes |v_2\rangle \dots \otimes |v_n\rangle$  where each  $v_i$  is a Haar random single-qubit state. Then there is an efficient randomized process with random coins  $r$ , that constructs the matrices  $P_i$  (depending on both  $r$  and  $|v\rangle$ ) such that the resulting state  $|\psi_{r,v}\rangle$  satisfies*

$$\mathbb{E}_{r,v} \langle \psi_{r,v} | H | \psi_{r,v} \rangle \geq \mathbb{E}_v \langle v | H | v \rangle + \Omega \left( \frac{\text{quad}(H)^2}{d|E|} \right).$$

*Proof.* Pick  $|v\rangle = \otimes_i |v_i\rangle$ , where each  $|v_i\rangle$  is chosen uniformly at random from Haar measure on qubits. Also choose  $n$  uniformly random real numbers  $\mu_i$  i.i.d in the interval  $[0, \frac{\pi}{2}]$ . The latter choice is made using the coins  $r$ . Given  $|v\rangle, r$  define  $P_i = e^{i\mu_i} |v_i\rangle \langle v_i^\perp| + e^{-i\mu_i} |v_i^\perp\rangle \langle v_i|$  (we drop the labels  $|v\rangle, r$  from  $P_i$  for convenience). Observe that  $\langle v_i | P_i | v_i \rangle = 0$  and  $\|P_i\| \leq 1$ , as required. Then  $\alpha_{v,r}$  (as given in Eq (5) ) can be evaluated to be

$$\begin{aligned} \alpha_{v,r} &= \mathbb{E}_{\{i,j\} \in E} \left| \left( e^{i(\mu_i + \mu_j)} \langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle - e^{-i(\mu_i + \mu_j)} \langle v_i, v_j | h_{ij} | v_i^\perp, v_j^\perp \rangle \right) \right| \\ &= 2 \mathbb{E}_{\{i,j\} \in E} \left| \text{Im} \left( e^{i(\mu_i + \mu_j)} \langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle \right) \right|. \end{aligned} \quad (\text{S9})$$

Let  $\langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle = e^{i\kappa_{i,j}} |\langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle|$  be the polar decomposition. Then

$$\left| \text{Im} \left( e^{i(\mu_i + \mu_j)} \langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle \right) \right| = |\sin(\mu_i + \mu_j + \kappa_{i,j})| \cdot |\langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle|.$$

Note that

$$\mathbb{E}_r |\sin(\mu_i + \mu_j + \kappa_{i,j})| = \frac{4}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} |\sin(\mu_i + \mu_j + \kappa_{i,j})| d\mu_i d\mu_j \geq \frac{2}{5},$$

for all  $\kappa_{i,j}$ . Then Eq S9 ensures that

$$\mathbb{E}_r \alpha_{v,r} = 2 \mathbb{E}_{\{i,j\} \in E} \mathbb{E}_r |\text{Im} \left( e^{i(\mu_i + \mu_j)} \langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle \right)| \geq \frac{4}{5} \cdot \mathbb{E}_{\{i,j\} \in E} |\langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle|.$$

Then we can evaluate

$$\begin{aligned} \mathbb{E}_{v,r} \alpha_{v,r} &\geq \frac{4}{5} \cdot \mathbb{E}_{\{i,j\} \in E} \int |\langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle| dv_i dv_j \\ &\geq \frac{4}{5} \cdot \mathbb{E}_{\{i,j\} \in E} \int |\langle v_i^\perp, v_j^\perp | h_{ij} | v_i, v_j \rangle|^2 dv_i dv_j \\ &= \frac{4}{5} \cdot \mathbb{E}_{\{i,j\} \in E} \int \text{tr} \left( |v_i^\perp, v_j^\perp\rangle \langle v_i^\perp, v_j^\perp| h_{ij} | v_i, v_j \rangle \langle v_i, v_j | h_{ij} \right) dv_i dv_j \\ &= \frac{4}{5} \cdot \mathbb{E}_{\{i,j\} \in E} \int \langle 11 | \left( U_i^\dagger \otimes V_j^\dagger \right) h_{ij} (U_i \otimes V_j) | 00 \rangle \langle 00 | \left( U_i^\dagger \otimes V_j^\dagger \right) h_{ij} (U_i \otimes V_j) | 11 \rangle dU_i dV_j \\ &= \frac{4}{5} \cdot \mathbb{E}_{\{i,j\} \in E} \int \langle 1100 | \left( U_{i1}^\dagger \otimes V_{j1}^\dagger \otimes U_{i2}^\dagger \otimes V_{j2}^\dagger \right) h_{i1,j1} \otimes h_{i2,j2} (U_{i1} \otimes V_{j1} \otimes U_{i2} \otimes V_{j2}) | 0011 \rangle dU_i dV_j, \end{aligned} \quad (\text{S10})$$

where in the second last equality we fixed a basis  $\{|0\rangle, |1\rangle\}$  for each qubit and introduced random unitaries  $U_i, V_j$  that specify  $|v_i\rangle = U_i|0\rangle, |v_j\rangle = V_j|0\rangle$ . Using the well known properties of Haar integral, we have

$$\begin{aligned} &\int \left( U_{i1}^\dagger \otimes V_{j1}^\dagger \otimes U_{i2}^\dagger \otimes V_{j2}^\dagger \right) h_{i1,j1} \otimes h_{i2,j2} (U_{i1} \otimes V_{j1} \otimes U_{i2} \otimes V_{j2}) dU_i dV_j \\ &= aI_{i1,i2} \otimes I_{j1,j2} + bS_{i1,i2} \otimes I_{j1,j2} + cI_{i1,i2} \otimes S_{j1,j2} + dS_{i1,i2} \otimes S_{j1,j2}. \end{aligned} \quad (\text{S11})$$

Above,  $I$  is the identity operator,  $S$  is the swap operator and the subscripts represent the qubits on which the operator acts. Coefficients  $a, b, c, d$  can be evaluated using the following system of equations, obtained from Eq. S11 by tracing each of the four operators.

$$\begin{aligned} (\text{tr}_{i,j} h_{ij})^2 &= 16a + 8b + 8c + 4d \\ \text{tr}_j (\text{tr}_i h_{ij} \text{tr}_i h_{ij}) &= 8a + 4b + 16c + 8d \\ \text{tr}_i (\text{tr}_j h_{ij} \text{tr}_j h_{ij}) &= 8a + 16b + 4c + 8d \\ \text{tr}_{i,j} (h_{ij}^2) &= 4a + 8b + 8c + 16d. \end{aligned}$$

One can solve for  $d$  to obtain

$$d = \frac{(\text{tr}_{i,j} h_{ij})^2}{36} + \frac{\text{tr}_{i,j} (h_{ij}^2)}{9} - \frac{\text{tr}_j (\text{tr}_i h_{ij} \text{tr}_i h_{ij}) + \text{tr}_i (\text{tr}_j h_{ij} \text{tr}_j h_{ij})}{18}.$$

In order to obtain a simpler lower bound and see that  $d$  is positive, we expand  $h_{ij} = \sum_{x,y} f_{x,y}^{i,j} \sigma_x^i \otimes \sigma_y^j$  in the two qubit Pauli basis. Then

$$\begin{aligned} (\text{tr}_{i,j} h_{ij})^2 &= 16 \left( f_{0,0}^{i,j} \right)^2, \quad \text{tr}_{i,j} (h_{ij}^2) = 4 \sum_{x,y} \left( f_{x,y}^{i,j} \right)^2, \\ \text{tr}_j (\text{tr}_i h_{ij} \text{tr}_i h_{ij}) &= 8 \sum_y \left( f_{0,y}^{i,j} \right)^2, \quad \text{tr}_i (\text{tr}_j h_{ij} \text{tr}_j h_{ij}) = 8 \sum_y \left( f_{y,0}^{i,j} \right)^2. \end{aligned}$$

Hence,

$$\begin{aligned} &\frac{(\text{tr}_{i,j} h_{ij})^2}{36} + \frac{\text{tr}_{i,j} (h_{ij}^2)}{9} - \frac{\text{tr}_j (\text{tr}_i h_{ij} \text{tr}_i h_{ij}) + \text{tr}_i (\text{tr}_j h_{ij} \text{tr}_j h_{ij})}{18} \\ &= \frac{4}{9} \left( \left( f_{0,0}^{i,j} \right)^2 + \sum_{x,y} \left( f_{x,y}^{i,j} \right)^2 - \sum_y \left( f_{0,y}^{i,j} \right)^2 - \sum_y \left( f_{y,0}^{i,j} \right)^2 \right) \\ &= \frac{4}{9} \left( \sum_{x>0,y>0} \left( f_{x,y}^{i,j} \right)^2 \right). \end{aligned}$$

Conjugating Eq. S11 with  $\langle 1100 | (\cdot) | 0011 \rangle$ , it can be seen that only the term corresponding to  $d$  survives and evaluates to 1. Thus, Eq. S10 gives

$$\begin{aligned} \mathbb{E}_{v,r} \alpha_{v,r} &\geq \\ \frac{4}{5} \mathbb{E}_{\{i,j\} \in E} &\left( \frac{(\text{tr}_{i,j} h_{ij})^2}{36} + \frac{\text{tr}_{i,j} (h_{ij}^2)}{9} - \frac{\text{tr}_j (\text{tr}_i h_{ij} \text{tr}_i h_{ij}) + \text{tr}_i (\text{tr}_j h_{ij} \text{tr}_j h_{ij})}{18} \right) \\ &= \frac{16}{45} \mathbb{E}_{\{i,j\} \in E} \left( \sum_{x>0,y>0} \left( f_{x,y}^{i,j} \right)^2 \right) = \frac{16}{45} \frac{\text{quad}(H)}{|E|}. \end{aligned}$$

Thus, using the convexity of square function,

$$\mathbb{E}_{v,r} \alpha_{v,r}^2 \geq \frac{16^2}{45^2} \left( \frac{\text{quad}(H)}{|E|} \right)^2 \geq \frac{1}{8} \left( \frac{\text{quad}(H)}{|E|} \right)^2.$$

This completes the proof by employing Theorem 2.  $\square$

### Triangle-free graphs

In this section we establish the second part of Theorem 4, which concerns triangle-free graphs. The proof is based on the following exact expression. It will be convenient in what follows to work in a local basis in which the product state of interest is  $|v\rangle = |0^n\rangle$ .

**Lemma 2 (Improvement for triangle-free Hamiltonians).** *Suppose  $G$  is a triangle-free, degree- $d$  graph. Suppose we are given single-qubit Hermitian operators  $\{P_i\}_{i \in [n]}$  satisfying  $P_i^2 = I$  and  $\langle 0|P_i|0\rangle = 0$  for all  $i \in [n]$ , and consider the state  $|\psi\rangle = e^{i \sum_{\{r,s\} \in E} \theta_{rs} P_r P_s} |0^n\rangle$  as a function of the real parameters  $\{\theta_{rs}\}$ . Define*

$$\alpha_{kl} = |\langle 00|[h_{kl}, P_k P_l]|00\rangle|$$

We can efficiently choose  $\theta_{ij} \in \{\pm\theta\}$  for each edge  $\{i, j\} \in E$  so that

$$\langle \psi|h_{kl}|\psi\rangle = \frac{1}{4}\text{Tr}(h_{kl}) + \frac{1}{4}\text{Tr}(h_{kl}Z_k Z_l) \cos^{2d-2}(2\theta) + \frac{1}{4}\text{Tr}(h_{kl}(Z_k + Z_l)) \cos^d(2\theta) + \frac{\alpha_{kl}}{2} \sin(2\theta) \cos^{d-1}(2\theta) \quad (\text{S12})$$

for all edges  $\{k, l\} \in E$

*Proof.* We have

$$\langle \psi|h_{kl}|\psi\rangle = \langle 0^n|V_{kl}^\dagger h_{kl}(\theta) V_{kl}|0^n\rangle \quad (\text{S13})$$

where  $h_{kl}(\theta_{kl}) = e^{-i\theta_{kl}P_k P_l} h_{kl} e^{i\theta_{kl}P_k P_l}$  and

$$V_{kl} = \prod_{\{k,s\} \in E \setminus \{k,l\}} e^{i\theta_{ks}P_k P_s} \prod_{\{r,l\} \in E \setminus \{k,l\}} e^{i\theta_{rl}P_r P_l} \quad (\text{S14})$$

$$= \prod_{\{k,s\} \in E \setminus \{k,l\}} (\cos(\theta) + i \sin(\theta_{ks})P_k P_s) \prod_{\{r,l\} \in E \setminus \{k,l\}} (\cos(\theta) + i \sin(\theta_{rl})P_r P_l). \quad (\text{S15})$$

Plugging Eq. (S15) into Eq. (S13) and using  $\langle 0|P_i|0\rangle = 0$  and the fact that  $G$  is triangle-free gives

$$\begin{aligned} \langle \psi|h_{kl}|\psi\rangle &= \sum_{A \subseteq N(k) \setminus \{l\}} \sum_{B \subseteq N(l) \setminus \{k\}} (\cos^2(\theta))^{2d-2-|A|-|B|} (\sin^2(\theta))^{|A|+|B|} \\ &\quad \cdot \langle 0^n | \left( \prod_{s \in A} P_k P_s \prod_{r \in B} P_r P_l \right) h_{kl}(\theta_{kl}) \left( \prod_{s \in A} P_k P_s \prod_{r \in B} P_r P_l \right) |0^n\rangle \end{aligned}$$

In the above we also used our choice  $|\theta_{ij}| = \theta$  for all edges  $\{i, j\} \in E$ . Observe that the matrix element appearing in the above depends only on the parity (even/odd) of  $|A|$  and  $|B|$ . In particular,

$$\langle \psi|h_{kl}|\psi\rangle = F_{EE} + F_{EO} + F_{OE} + F_{OO}$$

where the even/even term is

$$F_{EE} = \langle 00|h_{kl}(\theta_{kl})|00\rangle \left( \sum_{j=0,2,\dots} \binom{d-1}{j} (\cos^2(\theta))^{d-1-j} (\sin^2(\theta))^j \right)^2 \quad (\text{S16})$$

$$= \langle 00|h_{kl}(\theta_{kl})|00\rangle \frac{1}{4} (1 + \cos^{d-1}(2\theta))^2, \quad (\text{S17})$$

and by similar calculations one arrives at

$$F_{EO} = \langle 10|h_{kl}(\theta_{kl})|10\rangle \frac{1}{4} (1 - \cos^{2d-2}(2\theta)) \quad (\text{S18})$$

$$F_{OE} = \langle 01|h_{kl}(\theta_{kl})|01\rangle \frac{1}{4} (1 - \cos^{2d-2}(2\theta)) \quad (\text{S19})$$

$$F_{OO} = \langle 11|h_{kl}(\theta_{kl})|11\rangle \frac{1}{4} (1 - \cos^{d-1}(2\theta))^2 \quad (\text{S20})$$

Now for ease of presentation in the following we write  $c = \cos^{d-1}(2\theta)$  and  $a_{xy} = \langle xy|h_{kl}(\theta_{kl})|xy\rangle$ , for  $x, y \in \{0, 1\}$ . Then expanding the above expression gives

$$\langle \psi|h_{kl}|\psi\rangle = F_{EE} + F_{EO} + F_{OE} + F_{OO} \quad (\text{S21})$$

$$= a_{00} + (1-c) \left( \frac{a_{01} + a_{10}}{2} - a_{00} \right) + \frac{1}{4} (1-c)^2 (a_{11} + a_{00} - a_{01} - a_{10}). \quad (\text{S22})$$

Now let

$$b_{xy} = \langle xy|h_{kl}|xy\rangle. \quad (\text{S23})$$

So that

$$a_{00} = \cos^2(\theta)b_{00} + \sin^2(\theta)b_{11} + i \cos(\theta) \sin(\theta_{kl}) \langle 00|[h_{kl}, P_k P_l]|00\rangle.$$

We now fix the sign of  $\theta_{kl}$  so that

$$a_{00} = \cos^2(\theta)b_{00} + \sin^2(\theta)b_{11} + \frac{1}{2} \sin(2\theta)\alpha_{kl}. \quad (\text{S24})$$

With this choice we have

$$a_{11} = \cos^2(\theta)b_{11} + \sin^2(\theta)b_{00} - \frac{1}{2} \sin(2\theta)\alpha_{kl}. \quad (\text{S25})$$

To compute the third term in the above equation we used the fact that  $P_k^2 = P_l^2 = I$  and  $\langle 0|P_k|0\rangle = \langle 0|P_l|0\rangle = 0$  which implies

$$\langle 11|[P_k P_l, h_{kl}][11] \rangle = \langle 00|P_k P_l [P_k P_l, h_{kl}] P_k P_l |00\rangle = -\langle 00|[P_k P_l, h_{kl}][00] \rangle.$$

Similarly, by a direct calculation we get

$$a_{01} + a_{10} = b_{10} + b_{01}. \quad (\text{S26})$$

Plugging Eqs. (S24, S25, S26) into Eq. (S22) we get

$$\begin{aligned} \langle \psi|h_{kl}|\psi\rangle &= b_{00} + (a_{00} - b_{00})c + (1-c) \left( \frac{b_{01} + b_{10}}{2} - b_{00} \right) + \frac{1}{4}(1-c)^2 (b_{00} + b_{11} - b_{10} - b_{01}) \\ &= b_{00} + \frac{\alpha_{kl}}{2} \sin(2\theta) \cos^{d-1}(2\theta) + \sin^2(\theta) \cos^{d-1}(2\theta) (b_{11} - b_{00}) \\ &\quad + (1 - \cos^{d-1}(2\theta)) \left( \frac{b_{01} + b_{10}}{2} - b_{00} \right) + \frac{1}{4}(1 - \cos^{d-1}(2\theta))^2 (b_{00} + b_{11} - b_{10} - b_{01}) \end{aligned} \quad (\text{S27})$$

Rearranging the above expression we arrive at

$$\begin{aligned} \langle \psi|h_{kl}|\psi\rangle &= \frac{\alpha_{kl}}{2} \sin(2\theta) \cos^{d-1}(2\theta) + b_{00} \left( \frac{1}{4} + \frac{1}{4} \cos^{2d-2}(2\theta) + \frac{1}{2} \cos^d(2\theta) \right) \\ &\quad + b_{11} \left( \frac{1}{4} + \frac{1}{4} \cos^{2d-2}(2\theta) - \frac{1}{2} \cos^d(2\theta) \right) + (b_{01} + b_{10}) \left( \frac{1}{4} - \frac{1}{4} \cos^{2d-2}(2\theta) \right) \end{aligned} \quad (\text{S28})$$

By noting that  $\sum_{x,y} b_{xy} = \text{Tr}(h_{kl})$ ,  $\sum_{x,y} (-1)^{x+y} b_{xy} = \text{Tr}(h_{kl} Z_k Z_l)$ , and  $b_{00} - b_{11} = \frac{1}{2} \text{Tr}(h_{kl}(Z_k + Z_l))$ , we arrive at Eq. (S12). □

Using the expression in (S12), we prove the bound for triangle-free graphs from Theorem 4:

*Proof.* As shown above, the exact formula for the energy of  $|\psi\rangle = V(\vec{\theta})|0^n\rangle$  on a triangle-free graph is

$$\begin{aligned} \langle \psi|h_{kl}|\psi\rangle &= \frac{1}{4} \text{Tr}(h_{kl}) + \frac{1}{4} \text{Tr}(h_{kl} Z_k Z_l) \cos^{2d-2}(2\theta) \\ &\quad + \frac{1}{4} \text{Tr}(h_{kl}(Z_k + Z_l)) \cos^d(2\theta) + \frac{\alpha_{kl}}{2} \sin(2\theta) \cos^{d-1}(2\theta). \end{aligned} \quad (\text{S29})$$

Here  $\alpha_{kl}$  depends on the choices of  $P_k, P_l$ . We either choose  $P_i = X_i$  for all  $i$ , or  $P_i = (X + Y)_i / \sqrt{2}$  for all  $i$ , each with probability 1/2. Then

$$\mathbb{E}(\alpha_{kl}) = 2|\text{Re}(\langle 00|h_{kl}|11\rangle)| + 2|\text{Im}(\langle 00|h_{kl}|11\rangle)| \geq 2|\langle 00|h_{kl}|11\rangle|.$$

Substituting in Eq. (S29) gives

$$\begin{aligned} \mathbb{E}(\langle \psi | h_{kl} | \psi \rangle) &\geq \frac{1}{4} \text{Tr}(h_{kl}) + \frac{1}{4} \text{Tr}(h_{kl} Z_k Z_l) \cos^{2d-2}(2\theta) \\ &\quad + \frac{1}{4} \text{Tr}(h_{kl}(Z_k + Z_l)) \cos^d(2\theta) + |\langle 00 | h_{kl} | 11 \rangle| \sin(2\theta) \cos^{d-1}(2\theta). \end{aligned} \quad (\text{S30})$$

Now instead of using the starting state  $|0^n\rangle$ , suppose we start from a random computational basis state  $|s\rangle = X(s)|0^n\rangle$ . Running through the above argument in the rotated basis defined by  $|s\rangle$  we see that for a suitable random choice of  $\{P_i\}$  we have

$$\begin{aligned} \mathbb{E}(\langle \psi | h_{kl} | \psi \rangle) &\geq \frac{1}{4} (\text{Tr}(h_{kl})) + \frac{|\langle 00 | h_{kl} | 11 \rangle| + |\langle 01 | h_{kl} | 10 \rangle|}{2} \sin(2\theta) \cos^{d-1}(2\theta) \\ &\geq \frac{1}{4} (\text{Tr}(h_{kl})) + \frac{1}{4} |\text{Tr}(h_{kl} X_k X_l)| \sin(2\theta) \cos^{d-1}(2\theta). \end{aligned} \quad (\text{S31})$$

Here we used the fact that

$$\mathbb{E}_s(\text{Tr}(X(s)h_{kl}X(s)Z_kZ_l)) = \mathbb{E}_s(\text{Tr}(X(s)h_{kl}X(s)(Z_k + Z_l))) = 0.$$

Summing Eq. (S31) over all edges  $\{k, l\} \in E$  gives

$$\mathbb{E}(\langle \psi | H | \psi \rangle) \geq \frac{1}{4} (\text{Tr}(H)) + \frac{1}{4} \sin(2\theta) \cos^{d-1}(2\theta) \sum_{\{k,l\} \in E} |\text{Tr}(h_{kl} X_k X_l)|$$

Since there is nothing special about the  $X$ -basis we can again use our freedom to randomize the local basis of each qubit to get

$$\begin{aligned} \mathbb{E}(\langle \psi | H | \psi \rangle) &\geq \frac{1}{4} \text{Tr}(H) + \frac{1}{36} \sin(2\theta) \cos^{d-1}(2\theta) \sum_{\{k,l\} \in E} \sum_{Q,R \in \{X,Y,Z\}} |\text{Tr}(h_{kl} Q_k R_l)| \\ &\geq \frac{1}{4} \text{Tr}(H) + \frac{1}{36} \sin(2\theta) \cos^{d-1}(2\theta) \sum_{\{k,l\} \in E} \sum_{Q,R \in \{X,Y,Z\}} |\text{Tr}(h_{kl} Q_k R_l)|^2 / 4 \\ &= \frac{1}{4} \text{Tr}(H) + \sin(2\theta) \cos^{d-1}(2\theta) \frac{\text{quad}(H)}{36}, \end{aligned} \quad (\text{S32})$$

where in the second-to-last line we used the fact that  $|\text{Tr}(h_{kl} Q_k R_l)| \leq 4$  which follows from  $\|h_{kl}\| \leq 1$ . Finally, we can find the maximum value of the second term with respect to  $\theta$  by noting that  $\sin(2\theta) \cos^{d-1}(2\theta)$  reaches a maximum when  $\theta = \arcsin(\frac{1}{\sqrt{d}})$ . Using this fact we get

$$\mathbb{E}(\langle \psi | H | \psi \rangle) \geq \frac{1}{4} \text{Tr}(H) + \Omega\left(\frac{\text{quad}(H)}{\sqrt{d}}\right). \quad (\text{S33})$$

□

### IMPROVEMENT OF BOUNDED-DEPTH STATES

We prove Theorem 3. Given the  $d$ -regular graph  $G = (V, E)$ , we consider the state  $|v\rangle = W|0\rangle^n$ , where  $W$  has a maximum lightcone of size  $\ell$ . The aim is to increase the energy of  $|v\rangle$  with respect to  $H$ . The light cones of the edges have sizes at most  $2\ell$ . Define

$$F = \sum_{j=1}^n W|1\rangle\langle 1|_j W^\dagger. \quad (\text{S34})$$

The locality of  $F$  is  $\ell$ . Let  $A = i[H, F]$  and define  $|\psi\rangle = e^{iA\theta}|v\rangle$  (thus  $U = e^{iA\theta}$  in the statement of Theorem 3). We can write

$$A = \sum_{e \in E} i[h_e, F] := \sum_{e \in E} A_e, \quad (\text{S35})$$



where

$$A_e = i \sum_{j=1}^n [h_e, W|1\rangle\langle 1|_j W^\dagger] = i \sum_{j: \text{supp}(h_e) \cap \text{supp}(W|1\rangle\langle 1|_j W^\dagger) \neq \phi} [h_e, W|1\rangle\langle 1|_j W^\dagger]. \quad (\text{S36})$$

Any  $j$  satisfying  $\text{supp}(h_e) \cap \text{supp}(W|1\rangle\langle 1|_j W^\dagger) \neq \phi$  is in the light cone of  $h_e$ . Thus there are  $\leq 2\ell$  such  $j$ 's. For any such  $j$ ,  $W|1\rangle\langle 1|_j W^\dagger$  has locality  $\ell$ . Thus,  $A_e$  is supported on  $\leq 2\ell^2$  qubits. Further,

$$\|A_e\| \leq 2\ell \cdot \max_{j: \text{supp}(h_e) \subset \text{supp}(W|1\rangle\langle 1|_j W^\dagger)} \|[h_e, W|1\rangle\langle 1|_j W^\dagger]\| \leq 2\ell,$$

where we used

$$\|[h_e, W|1\rangle\langle 1|_j W^\dagger]\| = \|[h_e, W|1\rangle\langle 1|_j W^\dagger - I/2]\| \leq 2\|h_e\| \|W|1\rangle\langle 1|_j W^\dagger - I/2\| \leq 1.$$

We have

$$\langle \psi | H | \psi \rangle = \langle v | H | v \rangle - i \langle v | [A, H] | v \rangle \theta + \sum_{m=2}^{\infty} \frac{(-i\theta)^m}{m!} \langle v | [A, H]_m | v \rangle, \quad (\text{S37})$$

Now, using the identities  $F|v\rangle = 0$  and  $F \geq I - |v\rangle\langle v|$ , we find

$$-i \langle v | [A, H] | v \rangle \theta = \langle v | [[H, F], H] | v \rangle \theta = 2 \langle v | H F H | v \rangle \theta \geq 2\theta \langle v | H (I - |v\rangle\langle v|) H | v \rangle = 2\theta \text{Var}(H). \quad (\text{S38})$$

Thus, let us focus on the terms with  $m \geq 2$ . We upper bound

$$\langle v | [A, H]_m | v \rangle \leq \sum_{e \in E} \|[A, h_e]_m\| \leq |E| \max_e \|[A, h_e]_m\|. \quad (\text{S39})$$

Now, consider for each  $e$ ,

$$[A, h_e]_m = \sum_{e_1, \dots, e_m} [A_{e_m}, [A_{e_{m-1}} \dots [A_{e_1}, h_e]]], \quad (\text{S40})$$

where we used Eq. S35. Most terms are zero and we will bound the number of non-zero terms. We will use the following simple fact.

**Fact 1.** *Let  $S \subset V$ . The number of  $e$  such that the support of  $A_e$  overlaps with  $S$  is at most  $|S|\ell^2 d$ . For each such  $e$  and any operator  $O_S$  on  $S$ , the support of  $[O_S, A_e]$  is at most  $|S| + 2\ell^2$ .*

*Proof.* Since  $e$  is such that the support of  $A_e$  overlaps with  $S$ , there exist a  $j$  satisfying  $\text{supp}(h_e) \cap \text{supp}(W|1\rangle\langle 1|_j W^\dagger) \neq \phi$  for which the support of  $[h_e, W|1\rangle\langle 1|_j W^\dagger]$  overlaps with  $S$ . Thus, either  $\text{supp}(h_e) \cap S \neq \phi$  or  $j$  belongs to the light cone of  $S$ . Since the support of  $h_e$  overlaps with the light cone of  $j$  in the latter case, we have that the support of  $h_e$  overlaps with the light cone of the light cone of  $S$  (in both the cases). To upper bound the number of possible  $e$ , we hence we count the size of the light cone of the light cone of  $S$  ( $\leq |S|\ell^2$ ) and then count the number of edges intersecting with this light cone ( $\leq d|S|\ell^2$ ). For any such  $e$ , the support of  $[O_S, A_e]$  is contained in the union of  $S$  and the support of  $A_e$ . This completes the proof.  $\square$

Using Fact 1, let us estimate the number of  $(e_1, e_2, \dots, e_m)$  that contribute to Eq. S40. Setting  $S$  to be the set of two vertices of  $e$ , we find that the number of  $e_1$  is at most  $2d\ell^2$ . Arguing inductively, suppose we have fixed  $e_1, e_2, \dots, e_{k-1}$ . The support size of  $[A_{e_{k-1}}, [A_{e_{m-1}} \dots [A_{e_1}, h_e]]]$  is at most  $2 + 2(k-1)\ell^2$  (by Fact 1). Thus, the number of  $e_k$  contributing to Eq. S40 is at most  $(2 + 2(k-1)\ell^2)\ell^2 d$ . Hence, the total number of  $(e_1, \dots, e_m)$  is at most

$$(2d\ell^2) \cdot (2 + 2\ell^2)\ell^2 d \cdot (2 + 4\ell^2)\ell^2 d \dots (2 + 2(m-1)\ell^2)\ell^2 d \leq 2^m \cdot (m-1)! \cdot \ell^{2m-2} \cdot (2d\ell^2)^m \leq (m-1)!(4d\ell^4)^m.$$

Thus,

$$\begin{aligned} \|[A, h_e]_m\|_\infty &\leq (m-1)!(4d\ell^4)^m \max_{e_1, \dots, e_m} \|[A_{e_m}, [A_{e_{m-1}} \dots [A_{e_1}, h_e]]]\| \\ &\leq (m-1)!(4d\ell^4)^m \cdot 2^m \|h_e\| \cdot \max_{e_1, \dots, e_m} \|A_{e_1}\| \cdot \|A_{e_2}\| \dots \cdot \|A_{e_m}\| \\ &\stackrel{(a)}{\leq} (m-1)!(4d\ell^4)^m \cdot 2^m \cdot (2\ell)^m = (m-1)! \cdot (16d\ell^5)^m, \end{aligned} \quad (\text{S41})$$

where (a) uses  $\|A_e\| \leq 2\ell$ . Combining with Eq. S39, this ensures that

$$\begin{aligned} \sum_{m=2}^{\infty} \left| \frac{(i\theta)^m}{m!} \langle v | [[H, A]]_m | v \rangle \right| &\leq |E| \sum_{m=2}^{\infty} \frac{\theta^m}{m!} (m-1)! \cdot (16d\ell^5)^m \\ &\leq |E| \cdot \sum_{m=2}^{\infty} (16d\ell^5\theta)^m \\ &\leq 2|E| \cdot (16d\ell^5\theta)^2, \end{aligned} \quad (\text{S42})$$

where the last inequality assumes  $\theta \leq \frac{1}{32d\ell^5}$  (our choice below will satisfy this). Thus, using Eq. S37 and Eq. S38,

$$\begin{aligned} \langle \theta | H | \theta \rangle &\geq \langle v | H | v \rangle + 2\theta \text{Var}(H) - 2|E| \cdot (16d\ell^5\theta)^2 \\ &= \langle v | H | v \rangle + 2\theta |E| \left( \frac{\text{Var}(H)}{|E|} - 2(16d\ell^5)^2\theta \right). \end{aligned} \quad (\text{S43})$$

Setting  $\theta = \frac{\text{Var}(H)}{2^{10}d^2\ell^{10}|E|} \leq \frac{1}{32d\ell^5}$ , we conclude that

$$\langle \psi | H | \psi \rangle \geq \langle v | H | v \rangle + \frac{\text{Var}(H)^2}{2^{10}d^2\ell^{10}|E|}. \quad (\text{S44})$$

We highlight that the above proof can be applied with minor modifications to the more general case in which  $F$  is a  $\ell$ -local Hamiltonian with the unique ground state  $|v\rangle$  and constant spectral gap  $\gamma = \Omega(1)$ . In this case, we set the ground energy of  $F$  at 0, leading to the relations  $F|v\rangle = 0$  and  $F \succeq \gamma(I - |v\rangle\langle v|)$ . Thus, the first order contribution in (S38) is replaced by

$$-i\langle v | [A, H] | v \rangle \theta \geq 2\gamma\theta \text{Var}(H).$$

The higher order contributions are upper bounded in a manner similar to above.

## IMPROVEMENT FOR GENERAL $k$ -LOCAL HAMILTONIANS

Let  $G = (V, E)$  be a hypergraph with hyperedges of size at most  $k$  and  $n = |V|$  qubits on its vertices. We denote the number of hyperedges that contain  $i \in V$  by  $\text{deg}(i)$  and assume  $\text{deg}(i) \leq d$  for all  $i \in V$ . Consider a  $k$ -local Hamiltonian  $H = \sum_{R \in E} h_R$  where each local term  $h_R$  acts non-trivially on a subset  $R \in E$  of qubits with  $|R| \leq k$  and  $\|h_R\| \leq 1$ . Here without loss of generality we assume the input product states is  $|v\rangle = |0^n\rangle$ . We use a similar argument as in the proof of Theorem 1 to relate the improvement after applying an extension of the quantum circuit  $V(\theta)$  in Eq. (4) to the variance  $\text{Var}(H)$ . To this end, we write

$$\text{Var}(H) = \sum_{t=1}^k \langle 0^n | H Q_t H | 0^n \rangle$$

where  $Q_t$  are the projector onto the computational basis states with Hamming weight  $t$ . Note that operators  $Q_t$  with Hamming weight  $> k$  do not contribute. It holds that there exists a  $t$  such that

$$\langle 0^n | H Q_t H | 0^n \rangle \geq \frac{1}{k} \text{Var}(H).$$

Depending on  $t$ , our choice of circuit  $V(\theta)$  is a generalization of what we had before in Theorem 1. The set  $S$  contains all the collection of  $t$  different vertices  $\{j_1, j_2, \dots, j_t\}$  which fully reside in the support of at least one local term  $h_R$  of the Hamiltonian  $H$ . That is, there exists an  $R$  such that  $\{j_1, j_2, \dots, j_t\} \subseteq \text{supp}(h_R)$ . Define  $V(\theta) = e^{-i\theta L}$  where

$$L = \sum_{\{j_1, j_2, \dots, j_t\} \in S} (-1)^{a_{j_1, \dots, j_t}} P_{j_1} P_{j_2} \dots P_{j_t}.$$

Here each  $P_j$  is a single-qubit Pauli operator with the property  $\langle 0|P_j|0\rangle = 0$  that acts nontrivially only on qubit  $j$  and  $a_{j_1, \dots, j_t} \in \{0, 1\}$  is chosen so that

$$i\langle 0^n | (-1)^{a_{j_1, \dots, j_t}} [P_{j_1} P_{j_2} \dots P_{j_t}, H] | 0^n \rangle = |\langle 0^n | [P_{j_1} P_{j_2} \dots P_{j_t}, H] | 0^n \rangle|.$$

Let  $|\theta\rangle = e^{-i\theta L} |0^n\rangle$ . Then

$$\langle \theta | H | \theta \rangle = \langle 0^n | H | 0^n \rangle + \theta \sum_{\{j_1, \dots, j_t\} \in S} |\langle 0^n | [P_{j_1} P_{j_2} \dots P_{j_t}, H] | 0^n \rangle| + \text{Err},$$

where the higher order terms Err can be bounded as

$$\begin{aligned} |\text{Err}| &= \left| \sum_{m \geq 2} \frac{i^m \theta^m}{m!} \langle 0^n | [L, H]_m | 0^n \rangle \right| \\ &\leq |E| \sum_{m \geq 2} \frac{\theta^m}{m!} \left( 2kd \binom{k}{t-1} \right)^m \\ &\leq \left( 2kd \binom{k}{t-1} \right)^2 \theta^2 |E| e^{2kd \binom{k}{t-1} \theta}. \end{aligned} \quad (\text{S45})$$

In the second line, we used the fact that  $[L, H]_m = \sum_R [L, h_R]_m$  and each term  $[L, h_R]$  can be expanded as a sum of at most  $\left( kd \binom{k}{t-1} \right)^m$  non-zero terms each of norm at most  $2^m$ . This is because the operators  $P_{j_1}, P_{j_2}, \dots, P_{j_t}$  commute with each other for different choices of  $\{j_1, j_2, \dots, j_t\}$  and only those that overlap with the support of  $h_R$  may contribute. The number of such operators (i.e.  $|S \cap \text{supp}(h_R)|$ ) can be bounded by  $kd \binom{k}{t-1}$  as follows: There are at most  $k$  vertices in  $\text{supp}(h_R)$  and each vertex is in the support of  $\leq d$  other terms in the Hamiltonian. Given a vertex  $j \in \text{supp}(h_R)$  and an overlapping Hamiltonian term  $h_{R'}$  such that  $j \in \text{supp}(h_{R'})$ , there are  $\binom{k}{t-1}$  choices of vertices  $\{j_1, j_2, \dots, j_t\} \subseteq \text{supp}(h_{R'})$  that contain  $j$ . Hence, from the definition of set  $S$  follows that  $|S \cap \text{supp}(h_R)| \leq kd \binom{k}{t-1}$  (one can obtain tighter bounds using  $\langle 0 | P_j | 0 \rangle = 0$ ). Now define

$$\beta = \frac{1}{|E|} \sum_{\{j_1, \dots, j_t\} \in S} |\langle 0^n | [P_{j_1} P_{j_2} \dots P_{j_t}, H] | 0^n \rangle|. \quad (\text{S46})$$

It holds that  $\beta \leq 2 \binom{k}{t}$ . To see this, note that  $|\langle 0^n | [P_{j_1} P_{j_2} \dots P_{j_t}, H] | 0^n \rangle| \leq \sum_R |\langle 0^n | [P_{j_1} P_{j_2} \dots P_{j_t}, h_R] | 0^n \rangle|$ . Using the assumption  $\langle 0 | P_j | 0 \rangle = 0$ , it follows that the only choices of vertices  $\{j_1, \dots, j_t\}$  that may contribute in  $|\langle 0^n | [P_{j_1} P_{j_2} \dots P_{j_t}, h_R] | 0^n \rangle|$  are those which are fully contained in  $\text{supp}(h_R)$ . The number of such vertices is bounded by  $\binom{k}{t}$ . Using  $|\langle 0^n | [P_{j_1} P_{j_2} \dots P_{j_t}, h_R] | 0^n \rangle| \leq 2$ , we arrive at the claimed bound  $\beta \leq 2 \binom{k}{t}$ . We have

$$\langle \theta | H | \theta \rangle = \langle 0^n | H | 0^n \rangle + |E| \left( \theta \beta - \left( 2kd \binom{k}{t-1} \right)^2 \theta^2 e^{2kd \binom{k}{t-1} \theta} \right).$$

Choosing  $\theta = O\left(\frac{\beta}{k^2 d^2 \binom{k}{t-1}^2}\right)$  gives

$$\langle \theta | H | \theta \rangle \geq \langle 0^n | H | 0^n \rangle + \Omega\left(\frac{|E| \beta^2}{k^2 d^2 \binom{k}{t-1}^2}\right). \quad (\text{S47})$$

Now let  $\beta_1$  be given by Eq. (S46) with  $P_{j_1} P_{j_2} \dots P_{j_t} = X_{j_1} \otimes X_{j_2} \otimes \dots \otimes X_{j_t}$  for all  $\{j_1, \dots, j_t\} \in S$ . Define  $|\hat{e}_{j_1, \dots, j_t}\rangle = X_{j_1} \otimes X_{j_2} \otimes \dots \otimes X_{j_t} |0^n\rangle$  and the operator

$$p = \begin{pmatrix} 0 & e^{-i\frac{\pi}{2t}} \\ e^{i\frac{\pi}{2t}} & 0 \end{pmatrix}.$$

Let  $\beta_2$  be given by Eq. (S46) with  $P_{j_1} P_{j_2} \dots P_{j_t} = p_{j_1} \otimes p_{j_2} \otimes \dots \otimes p_{j_t}$  for all  $\{j_1, \dots, j_t\} \in S$ . Then, one can see that

$$\begin{aligned} \frac{\beta_1 + \beta_2}{2} &= \frac{1}{|E|} \sum_{\{j_1, \dots, j_t\} \in S} |\text{Im}(\langle \hat{e}_{j_1, \dots, j_t} | H | 0^n \rangle)| + |\text{Re}(\langle \hat{e}_{j_1, \dots, j_t} | H | 0^n \rangle)| \\ &\geq \frac{1}{|E|} \sum_{\{j_1, \dots, j_t\} \in S} d \cdot \left( \frac{|\text{Im}(\langle \hat{e}_{j_1, \dots, j_t} | H | 0^n \rangle)|^2}{d^2} + \frac{|\text{Re}(\langle \hat{e}_{j_1, \dots, j_t} | H | 0^n \rangle)|^2}{d^2} \right) \\ &= \frac{1}{d|E|} |\langle 0^n | H Q_t H | 0^n \rangle| \\ &\geq \frac{1}{kd|E|} \text{Var}(H). \end{aligned}$$

This implies either  $\beta_1$  or  $\beta_2$  is larger than  $\frac{1}{kd|E|} \text{Var}(H)$ . By plugging this into Eq. (S47) we arrive at

$$\langle \theta | H | \theta \rangle \geq \langle 0^n | H | 0^n \rangle + \Omega \left( \frac{\text{Var}(H)^2}{k^4 d^4 \binom{k}{t-1}^2 |E|} \right).$$

This bound is minimized by allowing  $t = \lceil k/2 + 1 \rceil$  which results in the following overall lower bound on the improvement to the energy of the input state  $|0^n\rangle$ :

$$\langle \theta | H | \theta \rangle \geq \langle 0^n | H | 0^n \rangle + \Omega \left( \frac{\text{Var}(H)^2}{2^{O(k)} d^4 |E|} \right). \quad (\text{S48})$$

We note that the  $\Omega(1/d^4)$  dependence of (S48) on the degree is quadratically worse than the bound we obtained for 2-local Hamiltonians in Theorem 1; It would be interesting to recover the  $\Omega(1/d^2)$  scaling in this case.

## LOCAL CLASSICAL ALGORITHMS

**Theorem 5.** Consider a two-local Hamiltonian  $H = \sum_{\{i,j\} \in E} h_{ij}$  where  $G = (V, E)$  is a  $d$ -regular triangle-free graph. There is an efficient randomized algorithm that computes a product state  $|v\rangle = \otimes_{i=1}^n |v_i\rangle$  satisfying

$$\mathbb{E}_v \langle v | H | v \rangle \geq \frac{1}{4} \text{Tr}(H) + \Omega \left( \frac{\text{quad}(H)}{\sqrt{d}} \right). \quad (\text{S49})$$

Note that in Eq. (S49) the first term on the right hand side is equal to the expected energy of  $H$  with respect to a random state  $\rho = I/2^n$ .

*Proof of Theorem 5.* It will be convenient to work in a local Pauli basis  $X, Y$  or  $Z$  chosen at random and independently for each qubit. We write  $h_{ij}$  in this randomly chosen product basis. Let us define  $w_{ij} = \text{Tr}(h_{ij})/4$  and

$$u_{ij}^x = \frac{1}{4} \text{Tr}(h_{ij} X_i X_j) \quad u_{ij}^y = \frac{1}{4} \text{Tr}(h_{ij} Y_i Y_j) \quad u_{ij}^z = \frac{1}{4} \text{Tr}(h_{ij} Z_i Z_j).$$

Due to the random choice of basis we have

$$\mathbb{E}[(u_{ij}^a)^2] = \frac{1}{9} \text{quad}(h_{ij}) \quad a \in \{x, y, z\}. \quad (\text{S50})$$

Following [S30, S31], we start with a random i.i.d assignment of pure product  $|v\rangle = \otimes_{i=1}^n |v_i\rangle$  states to the vertices. We then select a subset of vertices  $A$  uniformly at random. For any vertex  $i \in A$ , let  $N(i) = \{\{i, j\} \in E : j \notin A\}$  be the neighboring edges that contain exactly one vertex in  $A$  (i.e. vertex  $i$ ). The remaining edges that are not in  $\cup_{i \in A} N(i)$  either connect two vertices that are not in  $A$  or connect two vertices in  $A$ . We denote the former by  $M$  and the latter by  $M'$ .

The initial random pure state at each vertex  $\rho_i$  can be represented by  $\rho_i = \frac{1}{2}(I + r_i^x X_i + r_i^y Y_i + r_i^z Z_i)$ , where  $(r_i^x, r_i^y, r_i^z) \in \mathbb{R}^3$  is the Bloch vector with norm  $|r_i^x|^2 + |r_i^y|^2 + |r_i^z|^2 = 1$ . For a vertex  $i \in A$ , the total energy of the edges  $N(i)$  is given by

$$\sum_{j:\{i,j\} \in N(i)} \text{tr}[h_{ij}\rho_i \otimes \rho_j] = \sum_{j:\{i,j\} \in N(i)} w_{ij} + \sum_{j:\{i,j\} \in N(i)} (u_{ij}^x r_i^x r_j^x + u_{ij}^y r_i^y r_j^y + u_{ij}^z r_i^z r_j^z) + \sum_{j:\{i,j\} \in N(i)} D_{ij}(\vec{r}_i, \vec{r}_j) \quad (\text{S51})$$

where

$$D_{ij}(\vec{r}_i, \vec{r}_j) = \sum_{a \neq b} c_{ij}^{ab} r_i^a r_j^b + \sum_{a \in \{x,y,z\}} (d_{ij}^a r_i^a + e_{ij}^a r_j^a).$$

for some coefficients  $c_{ij}^{ab}, d_{ij}^a, e_{ij}^a$ . Using Cauchy–Schwarz inequality, we see that the first two terms in Eq. (S51) can be maximized by applying a local unitary on each vertex  $i \in A$  which rotates the state  $\rho_i$  to a state  $\tilde{\rho}_i$  with the Bloch vector

$$R_i^a = \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^a r_j^a \right) \left( \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^x r_j^x \right)^2 + \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^y r_j^y \right)^2 + \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^z r_j^z \right)^2 \right)^{-1/2} \quad (\text{S52})$$

for  $a \in \{x, y, z\}$ . When the denominator of Eq. (S52) is zero, the vector  $\vec{R}$  is chosen uniformly at random. Hence, we get

$$\begin{aligned} \sum_{j:\{i,j\} \in N(i)} \text{tr}[h_{ij}\tilde{\rho}_i \otimes \rho_j] &= \sum_{\{i,j\} \in N(i)} w_{ij} + \left( \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^x r_j^x \right)^2 + \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^y r_j^y \right)^2 + \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^z r_j^z \right)^2 \right)^{1/2} \\ &+ \sum_{j:\{i,j\} \in N(i)} D_{ij}(\vec{R}_i, \vec{r}_j), \end{aligned} \quad (\text{S53})$$

A property of this construction is that  $\mathbb{E}[R_i^a] = 0$  for  $a \in \{x, y, z\}$  and  $R_i, R_j$  are independent of each other for  $\{i, j\} \in M'$ . This follows from the triangle-freeness, the definition of the set  $N(i)$ , and the initial uniform i.i.d. distribution of the state of vertices. Moreover, we have

$$\mathbb{E}[r_j^a r_k^b] = \mathbb{E}[R_j^a R_k^b] = 0 \quad \text{whenever } a \neq b \quad .$$

Using these observations, the expected value of the total energy after  $\vec{R}$  the local improvements is

$$\begin{aligned} &\mathbb{E} \left[ \sum_{\{i,j\} \in M'} \text{tr}[h_{ij}\tilde{\rho}_i \otimes \tilde{\rho}_j] \right] + \mathbb{E} \left[ \sum_{\{i,j\} \in M} \text{tr}[h_{ij}\rho_i \otimes \rho_j] \right] + \mathbb{E} \left[ \sum_{i \in A} \sum_{j:\{i,j\} \in N(i)} \text{tr}[h_{ij}\tilde{\rho}_i \otimes \rho_j] \right] \\ &= \sum_{\{i,j\} \in E} w_{ij} + \mathbb{E} \left[ \sum_{i \in A} \left( \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^x r_j^x \right)^2 + \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^y r_j^y \right)^2 + \left( \sum_{j:\{i,j\} \in N(i)} u_{ij}^z r_j^z \right)^2 \right)^{1/2} \right] \\ &\geq \sum_{\{i,j\} \in E} w_{ij} + \mathbb{E} \left[ \sum_{i \in A} \left| \sum_{j:\{i,j\} \in N(i)} u_{ij}^z r_j^z \right| \right] \end{aligned} \quad (\text{S54})$$

The first term  $\sum_{\{i,j\} \in E} w_{ij}$  corresponds to the expected energy when the product states are chosen uniformly at random. The second term is a lower bound on the improvement achieved by the local updates which we now show is at least  $\Omega(|E|/\sqrt{d})$ .

For a fixed choice of the set  $A \subseteq V$ , we define the random variable  $\xi_i = \sum_{j:\{i,j\} \in N(i)} u_{ij}^z r_j^z$ . Using the ‘‘second moment method,’’ for  $t \in [0, 1]$ , we get

$$\Pr \left[ |\xi_i| \geq t \sqrt{\mathbb{E}[\xi_i^2]} \right] \geq (1 - t^2)^2 \frac{\mathbb{E}[\xi_i^2]^2}{\mathbb{E}[\xi_i^4]} \quad (\text{S55})$$

One way to sample uniform pure states over Bloch sphere is to uniformly draw  $\phi \sim [0, 2\pi]$ ,  $r_j^z \sim [-1, 1]$  and set  $r_j^x = \sqrt{1 - (r_j^z)^2} \cos \phi$  and  $r_j^y = \sqrt{1 - (r_j^z)^2} \sin \phi$ . Given this we have  $\mathbb{E}[r_j^z] = 0$ ,  $\mathbb{E}[(r_j^z)^2] = 1/3$ ,  $\mathbb{E}[(r_j^z)^3] = 0$ , and  $\mathbb{E}[(r_j^z)^4] = 1/5$ . Hence, using Corollary 9.6 of [S33], we have  $\mathbb{E}[\xi^4] \leq 9 \cdot \mathbb{E}[\xi^2]^2$ . Plugging this in (S55) implies that for a fixed choice of the set  $A$ , the expectation with respect to the random distribution of the initial product states for an arbitrary choice of  $t \in [0, 1]$  is

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{j:\{i,j\} \in N(i)} u_{ij}^z r_j^z \right| \right] &\geq \frac{1}{9} \cdot t(1-t^2)^2 \cdot \mathbb{E}[\xi^2]^{1/2} \\ &= \frac{1}{9} \cdot t(1-t^2)^2 \cdot \left( \sum_{j:\{i,j\} \in N(i)} (u_{ij}^z)^2 \mathbb{E}[(r_j^z)^2] \right)^{1/2} \\ &\geq \frac{1}{9\sqrt{3}} \cdot t(1-t^2)^2 \cdot \left( \sum_{j:\{i,j\} \in N(i)} (u_{ij}^z)^2 \right)^{1/2}. \end{aligned} \quad (\text{S56})$$

Finally, we calculate the expectation with respect to the set  $A \subseteq V$ . Note that the set  $N(i)$  is also a random variable determined by the set  $A$ . Conditioned on the event that the vertex  $i \in A$  and using Theorem 9.24 of [S33], we have

$$\Pr \left[ \sum_{j:\{i,j\} \in N(i)} (u_{ij}^z)^2 \geq \frac{1}{2} \sum_{j:\{i,j\} \in E} (u_{ij}^z)^2 \right] \geq \frac{1}{4e^2}.$$

Thus, we get

$$\begin{aligned} \mathbb{E} \left[ \sum_{i \in A} \left( \sum_{j:\{i,j\} \in N(i)} (u_{ij}^z)^2 \right)^{1/2} \right] &= \frac{1}{8\sqrt{2}e^2} \sum_{i \in V} \left( \sum_{j:\{i,j\} \in E} (u_{ij}^z)^2 \right)^{1/2} \\ &= \frac{1}{4\sqrt{2}e^2} \cdot \frac{1}{\sqrt{d}} \sum_{\{i,j\} \in E} (u_{ij}^z)^2. \end{aligned} \quad (\text{S57})$$

Finally, taking the expectation over the random choice of local basis and using Eq. (S50) we get

$$\begin{aligned} \mathbb{E} \left[ \sum_{i \in A} \left( \sum_{j:\{i,j\} \in N(i)} (u_{ij}^z)^2 \right)^{1/2} \right] &\geq \frac{1}{36\sqrt{2}e^2} \cdot \frac{1}{\sqrt{d}} \sum_{\{i,j\} \in E} \text{quad}(h_{ij}) \\ &\geq \frac{1}{36\sqrt{2}e^2} \cdot \frac{1}{\sqrt{d}} \text{quad}(H) \end{aligned} \quad (\text{S58})$$

We arrive at (S49) by plugging this into (S56) and using (S54).

□