A Short Path Quantum Algorithm for Exact Optimization

Matthew B. Hastings[1, 2]

[1]*Station Q, Microsoft Research, Santa Barbara, CA 93106-6105, USA*
[2]*Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA 98052, USA*

We give a quantum algorithm to exactly solve certain problems in combinatorial optimization, including weighted MAX-2-SAT as well as problems where the objective function is a weighted sum of products of Ising variables, all terms of the same degree D; this problem is called weighted MAX-ED-LIN2. We require that the optimal solution be unique for odd D and doubly degenerate for even D; however, we expect that the algorithm still works without this condition and we show how to reduce to the case without this assumption at the cost of an additional overhead. While the time required is still exponential, the algorithm provably outperforms Grover's algorithm assuming a mild condition on the number of low energy states of the target Hamiltonian. The detailed analysis of the runtime depends on a tradeoff between the number of such states and algorithm speed: having fewer such states allows a greater speedup. This leads to a natural hybrid algorithm that finds either an exact or approximate solution.

I. INTRODUCTION

While quantum algorithms are useful for many problems involving linear algebra (for example, finding eigenvalues[18], solving linear equations[15], and performing Hadamard transforms[1]), there are few proven speedups for combinatorial optimization problems. The most basic such speedup is Grover's algorithm[14], which gives a quadratic speedup over a brute-force search. For a problem such as finding the ground state of an Ising model on N spins, this can lead to a speedup from a brute force time $\mathcal{O}^*(2^N)$ (where $\mathcal{O}^*(\cdot)$ is big-O notation up to polylogarithmic factors, in this case polynomials in N) to $\mathcal{O}^*(2^{N/2})$. In the black box setting, Grover's algorithm is optimal[4], but for problems with a structure one might try to find a further speedup.

One attempt to find a speedup is the adiabatic algorithm[10]. Here, let H_Z be a Hamiltonian which is diagonal in the computational basis, such as

$$H_Z = \sum_{i,j} J_{i,j} Z_i Z_j, \tag{1}$$

where i labels different qubits and Z_i is the Pauli Z matrix on the i-th qubit. Then, consider the Hamiltonian

$$H_s = -(1-s)X + sH_Z, \tag{2}$$

where

$$X = \sum_i X_i \tag{3}$$

and X_i is the Pauli X matrix on the i-th qubit. At $s = 0$, the ground state of this Hamiltonian can be easily prepared. At $s = 1$, the ground state is the ground state of H_Z. If the spectral gap of H_s between the ground and first excited state is only polynomially small for $s \in [0, 1]$, then one may adiabatically evolve the ground state from $s = 0$ to $s = 1$ in polynomial time.

Unfortunately, the gap may become superpolynomially small. Indeed, it was argued using ideas from Anderson localization[2] that the gap may become as small as $N^{-\text{const.} \times N}$ so that the time required for adiabatic evolution is of order $N^{\text{const.} \times N}$ which is much slower than even classical brute force search. While the actual behavior for random instances may be more complicated than this[20], specific examples[33] can show this behavior.

This is perhaps not surprising. The problem of finding the ground state of H_Z is extremely hard, even if we restrict $J_{ij} \in \{-1, 0, 1\}$ for all i, j. While some speedups are known for bounded degree (here, we consider the graph with spins as vertices and an edge between vertices if the corresponding element of J is nonzero, and degree refers to the degree of this graph)[11,12,16,31], if we only allow polynomial space then the fastest classical algorithms for arbitrary J take a time $\mathcal{O}^*(2^N)$. If we allow exponential space then it is possible to reduce this time to $\mathcal{O}^*(2^{\omega N/3})$, where ω is the matrix multiplication exponent[34]. However, not only does that algorithm require exponential space, but it is not known how to give a Grover speedup of this algorithm, so that no quantum algorithm is known taking time $\mathcal{O}^*(2^{cN/2})$ for any $c < 1$. Finally, this algorithm is specific to constraint satisfaction problems where each constraint only involves a pair of variables, rather than a triple or more.

Here we present a quantum algorithm to find the ground state of H_Z which improves on Grover's algorithm in many cases. In the next subsection we give the problem definition and define which H_Z we consider.

A. Problem Definition

The problem is to find the ground state of H_Z (i.e., eigenstate of H_Z with minimum eigenvalue; eigenvalues of H_Z, H_s will often be called "energies") assuming that the ground state energy E_0 is known. We impose the following conditions on H_Z. Let N be the number of qubits. We let H_Z be any Hamiltonian that is a weighted sum of products of Pauli Z operators, with each product containing exactly D such operators on distinct qubits for some given D. That is, the case $D = 2$ is the Ising model, the case $D = 4$ is a sum of terms $Z_i Z_j Z_k Z_l$ for i, j, k, l all distinct, and so on. We take $D = \mathcal{O}(1)$. We emphasize that all products must have the same D so that we do not allow for example $H_Z = Z_1 Z_2 + Z_1 Z_2 Z_3$; this is for a reason explained later.

Each product has a weight that is an integer. We define J_{tot} to be the sum of the absolute values of the weights. We require that $J_{tot} = \mathcal{O}(\text{poly}(N))$. That is, we fix some $\beta > 0$ and require $J_{tot} = \mathcal{O}(N^\beta)$. (If all weights are chosen from $\{-1, +1\}$ we have $J_{tot} = \mathcal{O}(N^2)$.)

We require that H_Z has a unique ground state for D odd while for D even we require that H_Z has a doubly degenerate ground state. For even D, the operator $\prod_i X_i$ commutes with the Hamiltonian, flipping all spins, so that every eigenvalue has an even degeneracy. We call this assumption the "degeneracy assumption". We will analyze the performance of the algorithm without this assumption in future work. In an Appendix we show how to reduce to the case without this assumption at the cost of an additional time overhead.

We give two theorems 1,2 which describe the performance of the algorithm; the different theorems correspond to different choices of the parameters in the algorithm. Both theorems show, roughly, that at least one of two things holds: the algorithm finds the ground state in a certain expected time or H_Z has a large number of low energy states, i.e., eigenstates with eigenvalue close to E_0.

These theorems can be applied in one of two ways. One way is to define a promise problem, in which we are promised that H_Z does *not* have such a large number of low energy states as well as promised the degeneracy assumption on H_0. Then, the algorithm solves this promise problem. For this promise problem, while the algorithm is given E_0, it is not necessary to know E_0 in advance since one can try all possible E_0 with only polynomial overhead.

Alternatively, we can consider H_Z without such a promise on the number of low energy states but still including the promise on the degeneracy assumption. Then, we give a hybrid algorithm which tries both running the quantum algorithm here to find the exact ground state as well as random sampling (or a Grover search) to find an approximate ground state. This hybrid algorithm will be explained after theorem 1.

The problem of minimizing H_Z is also known as MAX-ED-LIN2, as we assume that each term is exactly of degree D, rather than having degree at most D; of course, there is an overall sign difference also, as we try to minimize H_Z rather than maximizing. Due to the uniqueness conditions on H_Z, we call this problem UNIQUE-MAX-ED-LIN2.

However, we can reduce any instance of MAX-2-LIN2 with a unique solution to UNIQUE-MAX-E2-LIN2, where MAX-2-LIN2 allows terms to have degree 1 or 2. Up to a sign difference, MAX-2-LIN2 with a unique solution means that we consider $H_Z = \sum_{i,j} J_{i,j} Z_i Z_j + h_i Z_i$, with integer $J_{i,j}, h_i$ and $\sum_{i,j} |J_{ij}| + \sum_i |h_i| = J_{tot}$, with H_Z required to have a unique ground state. We can find the ground state of this problem by finding one of the two ground states of a problem on $N + 1$ spins defined by $H_Z = \sum_{i,j} J_{i,j} Z_i Z_j + \sum_i h_i Z_0 Z_i$, where 0 is the added spin.

B. Main Results

We now give the main results. The parameters B, K enter into the definition of the algorithm, given later; the algorithm makes use of a family of Hamiltonians $H_s = H_Z - sB(X/N)^K$ so that B sets the strength of a term that is proportional to a power (set by K) of the transverse field. In each theorem, the statement about the number of low energy states of H_Z is first expressed as a statement about probability distributions with high entropy and low energy which then implies the statement about the number of low energy states.

We give two distinct theorems. The first theorem considers the case that $K = C \log(N)$ for $C > 0$ a fixed constant. The second theorem is more general, and allows K to be an arbitrary odd integer with $K \geq 3$; K may depend on N and the first theorem follows from the second. The first theorem is simply given as a way to explain some of the results in the particular case of a logarithmically growing K, when some of the results simplify; the second theorem gives stronger results but they are more complicated. Even in the case of a logarithmically growing K, the second theorem is stronger.

Fixed K, compared to logarithmically growing K, leads to a larger speedup but requires a stronger promise.

In an abuse of notation, when we write $K = C \log(N)$, we mean that K is taken to be the smallest odd integer larger than $C \log(N)$. We explicitly write the dependence on these constants in the equations, not hiding them in big-O notation. The quantity b is a real number, $0 < b < 1$. We take $b < 1$ else the second possibility (item 2 in the list) is trivially true and does not imply any interesting constraint on the density of states.

We let $W(E)$ denote the number of computational basis states with expectation value E for H_Z.

All entropies are measured in bits rather than nats. All logarithms are to base 2 unless otherwise specified.

Theorem 1. *Assume that H_Z obeys the degeneracy assumption. Suppose that $B = -bE_0$ and $K = C\log(N)$. Then, at least one of the following holds:*

1. *The algorithm finds the ground state in expected time*

$$\mathcal{O}^*\left(2^{N/2}\exp\left[-\frac{b}{2CD}\frac{N}{\log(N)}\right]\right).$$

2. *There is some probability distribution $p(u)$ on computational basis states with entropy at least*

$$S^{(comp)} \geq N \cdot (1 - \mathcal{O}(1)/C)$$

   and with expected value of H_Z at most $(1-b)E_0 + \mathcal{O}(1) \cdot \frac{J_{tot}}{N^2}C^2D^2\log(N)^2$. Further, for any $\eta > 0$, for some

$$E \leq E_0 + (1+\eta)\left(b|E_0| + \mathcal{O}(1) \cdot \frac{J_{tot}}{N^2}C^2D^2\log(N)^2\right),$$

   we have

$$\log(W(E)) \geq N \cdot (1 - \mathcal{O}(1)\frac{1+\eta}{\eta}\frac{1}{C}) - \frac{1+\eta}{\eta}\mathcal{O}(\log(N)).$$

Theorem 1 implies the following corollary about a hybrid algorithm:

Corollary 1. *There is an algorithm that, given H_Z and E_0, either outputs "approximate" or "exact". If it outputs "approximate", it returns also a state with energy at most $E_{approx} \equiv E_0 + (1.01)\left(b|E_0| + \mathcal{O}(1) \cdot \frac{J_{tot}}{N^2}C^2D^2\log(N)^2\right)$ and takes time at most $\mathcal{O}^*(2^{\mathcal{O}(1)N/C})$. If it outputs "exact", it outputs an exact ground state of H_Z. The expected run time of the algorithm (averaged over both approximate and exact outputs, rather than conditioned on an output) is $\mathcal{O}^*\left(2^{N/2}\exp\left[-\frac{b}{2CD}\frac{N}{\log(N)}\right]\right).$*

Proof. Take $\eta = 0.01$.

Run the following algorithm. First, try repeated random sampling of states to find a state with energy at most E_{approx}. Take a total of n_{samp} samples. If any sample succeeds, terminate the algorithm, returning "approximate" and the given state. If item 2 of the theorem holds, each sample succeeds in finding such a state with probability at least $2^{-\mathcal{O}(1)N/C - \mathcal{O}(\log(N))} = 2^{-\mathcal{O}(1)N/C}/\text{poly}(N)$, so one can choose $n_{samp} = \mathcal{O}^*(2^{\mathcal{O}(1)N/C})$ such samples and succeed with probability at least $1 - 2^{-N}$.

If no sample succeeds, then run the quantum algorithm of this paper in parallel with a brute force search, until one of them finds an exact ground state, returning "exact" and the state found.

The run time bound on the approximate output holds by construction. We now show the expected run time bounds. If item 1 holds, then the algorithm of this paper succeeds in expected time $\mathcal{O}^*\left(2^{N/2}\exp\left[-\frac{b}{2CD}\frac{N}{\log(N)}\right]\right)$ so the expected run time bound holds. If item 2 holds and the repeated random sampling fails to find an approximate state then the brute force search will find the ground state in time $\mathcal{O}^*(2^N)$; since the probability that the repeated random sampling fails is $\leq 2^{-N}$, this adds a negligible amount to the expected time of the algorithm.

Remark: the repeated random sampling can be quadratically improved with Grover search but this only improves constants which are hidden in the big-O notation. $\square$

We now give the case of arbitrary K. The function $\tau(\cdot)$ in theorem 2 is a continuous increasing function, taking $[0,1]$ to $[0,1]$. It is defined in lemma 11. It is differentiable on this interval, except at zero; for small σ,

$$\tau(\sigma) = \Theta(\sqrt{\frac{\sigma}{-\log(\sigma)}}). \tag{4}$$

In Fig. 1 we plot $\tau(\cdot)$.

Theorem 2. *Assume that H_Z obeys the degeneracy assumption. Suppose that $B = -bE_0$ with $0 \leq b < 1$. Suppose that $K \geq 3$ and $K > \beta$, with K odd. Then, at least one of the following holds:*
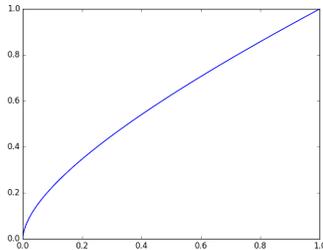
FIG. 1: A plot of $\tau(\cdot)$.

1. The algorithm finds the ground state in expected time

$$\mathcal{O}^*\Big(2^{N/2}\exp\Big[-\frac{b}{2DK}N\Big]\Big).$$

2. For some $X_0 \geq X_{min} = N \cdot (4B)^{-1/K}$, there is some probability distribution $p(u)$ on computational basis states with entropy in bits at least

$$S^{(comp)} \geq N\tau^{-1}\Big(\frac{X_0 - X_{min}/K}{N}\Big)$$

and with expected value of H_Z at most

$$E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B\Big(\frac{X_0 + X_{min}/K}{N}\Big)^K \cdot \mathcal{O}(1).$$

Further, for some function $F(S)$ with

$$F(S) = E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B(\tau(S/N))^K \cdot \mathcal{O}(1), \tag{5}$$

then for some $E > E_0$ we have $\log(W(E)) \geq F^{-1}(E) - \mathcal{O}(\log(N))$.

There are two parameters b, K that can be adjusted in this algorithm, so the implications of this theorem take some unpacking. We now argue, however, that for most problems for many choices of b, K the conditions of items 2. of the theorem will not be satisfied and hence a nontrivial speedup will occur from item 1. The next paragraph will deliberately be more heuristic, and will draw on some physics intuition, since the goal is not to prove a result but rather to argue that certain conditions do not occur in most examples.

Let us consider a simple example, $K = 99, b = 1/99$. Then, either the algorithm gives a speedup over a Grover search, taking expected time $\mathcal{O}^*(2^{cN/2})$ for a c slightly less than 1, or there are a large number of computational basis states with low energy for H_Z. Let us see whether we can satisfy $\log(W(E)) \geq F^{-1}(E)$ for any $E > E_0$. For E near 0, we have $\log(W(E))$ close to N since a typical state has energy E near 0. If we had taken $b = 1$, we could then satisfy $N \geq F^{-1}(0)$. However, the function $F(S)$ is bounded at $E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B\mathcal{O}(1)$, and the constant hidden in the big-O notation $\mathcal{O}(1)$ is such that for $b = 1/99$ we have $F(S)$ bounded away from zero. Indeed, $F(S)$ will be upper bounded by roughly $(1 - b\mathcal{O}(1))E_0$. Let us, for clarity in explanation pretend that the constant hidden in the $\mathcal{O}(1)$ is equal to 1. Then, we could have a problem if $F^{-1}(\frac{98}{99}E_0) \approx N$. However, this means an enormous (entropy almost equal to N) number of states have energy which is $98/99$ times the ground state energy. Now let us instead ask whether the conditions of item 2. can be satisfied for X_0 close to 0. Indeed, for an example such as the Sherrington-Kirkpatrick spin glass[17], the entropy $\log(W(E))$ has a power law dependence on $E - E_0$, with the scaling behavior well-understood[9,25] from the Parisi solution[26,27] giving $\log(W(E)) \sim N \cdot ((E - E_0)/|E_0|)^{2/3}$. By choosing K large, from studying the limiting behavior of c we find that we might satisfy the condition at small $E - E_0$ if $\log(W(E)) \sim N \cdot ((E - E_0)/|E_0|)^{2/99}$ up to some polylogs in energy. So, again, we would not expect to obey item 2. at small X_0 since $2/99 < 2/3$. Note also that for $K = 99$, since $B \leq \mathcal{O}(N^2)$ we have $X_{min} = \Omega(N^{1-2/99}) = \Omega(N^{97/99})$ so that X_{min} cannot be too small.

On the other hand, a model such as the "toy model" of Ref. 33 has a unique ground state but has $\log(W(E)) \approx N/2$ for $E = E_0 + 1$. Thus, for this model, we cannot consider fixed K without satisfying item 2. of the theorem but we can still obtain a nontrivial speedup by taking $K = C\log(N)$, albeit not giving $\mathcal{O}^*(2^{cN/2})$ with $c < 1$.

Even in the case $K = C \log(N)$, theorem 2 gives tighter bounds than theorem 1. The entropy in theorem 1 is the entropy for $X_0 = X_{min}$, i.e., it is the minimum possible, while the energy given in theorem 1 is the largest possible that can occur from theorem 2 for $K = C \log(N)$ and this value of energy would not occur for $X_0 = X_{min}$. Thus, theorem 2 gives stronger results involving a functional relationship between entropy and energy.

C. Outline

The algorithm has three distinct ideas. First we use a short path evolution, where the initial state is not a ground state of the initial Hamiltonian. That is, rather than trying to follow the full evolution from a transverse field Hamiltonian that is diagonal in the X basis to H_Z which is diagonal in the Z basis, we instead follow only a "short path", keeping a term proportional to H_Z in the Hamiltonian fixed and varying a transverse field-like term from a small value to zero. We called this term "transverse field-like" rather than a transverse field, because we add a more complicated term $-B(X/N)^K$ that is proportional to a power of the transverse field. This is the second idea and plays a key role in avoiding small spectral gaps. The third idea uses measurements to obtain adiabatic evolution with exponentially small error; similar ideas are in Ref. 28,32.

We will show that either the algorithm succeeds in giving a nontrivial speedup or a certain spectral gap becomes small. However, for such a gap to become small, there must be some state with a large expectation value for $-B(X/N)^K$ and with a small expectation value for H_Z. Roughly, the idea then is that for a large K, this imposes strong constraints on the expectation value of X. For example, if the state is an eigenstate of X, then we would have $\langle X/N \rangle = \langle (X/N)^K \rangle^{1/K}$, and so for large K, even a small expectation value of $(X/N)^K$ would give a large expectation value for X/N. Technically, the implementation is more complicated than this because there may be large fluctuations in X in the state. Ignoring these technical details for the moment, the next idea is that given a large expectation value of X, we use log-Sobolev estimates on entropy to relate the expectation value of X to the entropy in the computational basis. This is then used to show that there must be many eigenstates of H_Z with low energy.

Other technical tools used include a Brillouin-Wigner perturbation theory to compute overlaps and a method of localizing wavefunctions in X; this method is used to deal with the problem of large fluctuations in X.

In sections II,III we give the algorithm. In section IV we explain the Brillouin-Wigner perturbation theory. We then consider two cases, depending on the the spectrum of the Hamiltonian QH_sQ, where Q projects onto the states of H_Z with energy greater than E_0. Theorem 3 in section V assumes the first case and gives bounds on the spectral gap of H_s and shows the speedup of the algorithm. Theorem 4 in section VI assumes the second case and shows the existence of probability distributions over computational basis states with high entropy and low energy. Putting these results together, theorems 1,2 follow as shown in section VII; there, we use the high entropy and low energy probability distributions to lower bound $W(E)$. In section VIII, we briefly discuss some other results and also remark on a hybrid algorithm.

Section VI will use the ground state degeneracy assumption for convenience, but we will show how to remove this assumption for the results of this section. Section V will rely more heavily on the ground state degeneracy assumption; we will remove this assumption in future work using a modified perturbation theory.

II. THE SHORT PATH ALGORITHM

The algorithm that we use is based on applying amplitude amplification[8] to another algorithm. This latter algorithm is Algorithm 1. Remark: The case $K = 1, B = cN$ corresponds to a transverse magnetic field of strength c. Also, note the direction of evolution from $s = 1$ to $s = 0$.

To analyze this algorithm for even D we need a definition. For D even, we define the "even subspace" to be the eigenspace of $\prod_i X_i$ with eigenvalue $+1$ and define the "odd subspace" to be the eigenspace of $\prod_i X_i$ with eigenvalue -1. The Hamiltonian H_s commutes with $\prod_i X_i$ and ψ_+ is in the even subspace; the measurement algorithm will be chosen to preserve the eigenvalue of $\prod_i X_i$. When we analyze the algorithm for even D, all references for the remainder of the paper to the Hamiltonian will refer to the Hamiltonian projected into the even subspace and all vectors will be in the even subspace. We will remark on this as needed.

At $s = 0$, H_s has a unique ground state; call this state $|0\rangle$; for even D this state $|0\rangle$ is the unique ground state of H_Z in the even subspace and it is an equal amplitude superposition of two computational basis states. Let $\psi_{0,s}$ be the ground state of H_s. By Perron-Frobenius this state is unique for $s \geq 0$ for K odd; to see this, note that one can flip any given spin K times to attain a single spin flip so that the matrix H_s is irreducible in the computational basis, i.e., consider the term X_i^K for any given i. When we refer to uniqueness of a state, we mean uniqueness up to an overall phase.

Algorithm 1 Short-Path (unamplified version)

---

1. Prepare the wavefunction in the state $\psi_+ = |+\rangle^{\otimes N}$.

2. Use the measurement algorithm of section III to evolve under the Hamiltonian H_s from $s = 1$ to $s = 0$. where

$$H_s = H_Z - sB(X/N)^K, \tag{6}$$

where

$$X = \sum_i X_i. \tag{7}$$

where K is a positive odd integer and B is a scalar chosen later.

When we refer to a Hamiltonian H_s, it will be assumed that $s \in [0,1]$ unless explicitly stated otherwise. We pick $B > 0$ so that H_s has all off-diagonal entries non-positive in the computational basis.

3. Measure the state in the computational basis and compute the value of H_Z after measuring. If this value is equal to E_0 then declare success and output the computational basis state.

We will compute the squared overlap

$$P_{ov} \equiv |\langle \psi_+ | \psi_{0,1} \rangle|^2. \tag{8}$$

later. All state vectors that we write will be assumed to be normalized to have unit norm, except where explicitly stated otherwise later.

Suppose that the gap remains $\Omega(1/\text{poly}(N))$ along the path. Indeed, later we will give conditions under which the gap is $\Omega(1)$. Then, for any $P_{succ} < 1, \epsilon > 0$ we can use the measurement algorithm explained in section III to produce, with probability at least $P_{ov}P_{succ}$, a state equal to $\psi_{0,0}$ up to some error ϵ. This algorithm takes a time that is $\mathcal{O}(\text{poly}(N, (1 - P_{succ})^{-1}), \log(1/\epsilon))$. This measurement algorithm has the advantage, compared to adiabatic evolution, that we can achieve a better scaling with the error ϵ.

Thus, the quantum algorithm described above succeeds with probability at least $P_{ov}P_{succ} - \epsilon$ in finding the ground state of H_Z. We take $\epsilon = 2^{-N}$ so that the error ϵ is negligible compared to P_{ov} computed later. We choose $P_{succ} = 1/2$. Hence, applying the method of amplitude amplification to the evolution, one obtains an algorithm which succeeds in producing the ground state of H_Z in expected time $\mathcal{O}(P_{ov}^{-1/2} P_{succ}^{-1/2} \text{poly}(N, \log(1/\epsilon)))$.

III. MEASUREMENT ALGORITHM FOR ADIABATIC EVOLUTION WITH SMALLER ERROR

We now explain how to do step **2.** of algorithm 1. We explain the method in a general setting and then show that in the particular case here, it can be reduced to a single measurement.

Suppose that we have a path of Hamiltonians H_s, for $s \in [0,1]$. Suppose further that all of these Hamiltonians have a unique ground state $\psi_{0,s}$, separated from the rest of the spectrum by a gap at least $\Delta = \Omega(1/\text{poly}(N))$ and suppose that $\|\partial_s H_s\| = \mathcal{O}(\text{poly}(N))$. Suppose that for any s, we can simulate time evolution under H_s for time t up to error δ in a time that is $\mathcal{O}(\text{poly}(N, t, \log(1/\delta)))$. This holds for the Hamiltonians H_s considered above so long as $B = \mathcal{O}(\text{poly}(N))$ using any of several different algorithms in the literature which achieve this time[5–7,22,23].

We seek a quantum algorithm that, taking a state ψ as input with overlap $P_{ov} = |\langle \psi | \psi_{0,1} \rangle|^2$, will succeed, with probability at least $P_{ov}P_{succ}$ for P_{succ} close to 1, in giving an output state which is equal to $\psi_{0,0}$ up to some error ϵ, with ϵ exponentially small. A natural way to do this is to follow adiabatic evolution of the Hamiltonians, i.e., to simulate a time-dependent Hamiltonian which slowly changes from H_1 to H_0, taking ψ as input to the evolution and $\tilde{\psi}$ as output. Unfortunately, this method has two problems, one minor and one major. We will explain these problems before giving a different approach using measurements that avoids these problems.

The minor problem is that we must then simulate time-dependent evolution of a Hamiltonian; this problem is not too serious as for example the Taylor series approach[7] can simulate this evolution in a time that is $\mathcal{O}(\text{poly}(N, t, \log(1/\delta)))$ even for slowly varying Hamiltonians. The major problem is that there may be diabatic transitions from the ground state $\psi_{0,s}$ to some excited state along the path. While there is some controversy about the exact error estimates in the adiabatic theorem[3,29], even the best estimates give an error that is super-polynomially small but not exponentially small in the evolution time. Hence, even taking these estimates, in order to obtain an error δ that is exponentially small in N as we will need, we need a super-polynomial evolution time.

It is possible that a careful analysis of the error terms in the adiabatic algorithm would show that the required evolution time is not in fact too large. However, given some question about even simpler error estimates in the adiabatic theorem, we prefer not to use this method.

Instead we use the following algorithm 2. Later we will show that for the specific H_s considered in this paper, it suffices to use the simpler algorithm 3 given later. We present algorithm 2 here because it also works for more general paths of Hamiltonians H_s, so long as a spectral gap is present and the adiabatic evolution can be simulated.

Algorithm 2 Measurement Algorithm

---

1. Let ψ be the input state.

2. Phase estimate ψ using H_1. If the energy estimate is larger than $E_{0,1} + \Delta/2$, then terminate the algorithm and return failure. Else continue.

3. Adiabatically evolve ψ from H_1 to H_0.

4. Phase estimate ψ using H_1. If the energy estimate is larger than $E_{0,1} + \Delta/2$, then terminate the algorithm and return failure. Else declare success and return ψ.

One can also replace the adiabatic evolution of step **3.** with a sequence of measurements as in Ref. 28. This has the advantage that one does not need to simulate a time dependent evolution.

In algorithm 2, we use the initial phase estimation to approximately project onto $\psi_{0,1}$. Then we adiabatically evolve. Finally we repeat with another phase estimation to project onto $\psi_{0,0}$. We now consider the error in the phase estimation. We consider the phase estimation scheme of Ref. 19 though others are possible. There are two ways to quantify the error. One is the probability of error and the other is the precision. For us, it suffices to have a precision that is smaller than $\Delta/2$ so that one distinguishes the energy of the ground and first excited states. We take the probability of error in any step to be ϵ so that the total probability of error is $\mathcal{O}(\epsilon)$. If controlled time evolution can be implemented exactly, the phase estimation has an overhead that is only logarithmic in the error. For constant K, we can use any of the time evolution algorithms above or use[28] to implement the controlled time evolution with error $\mathcal{O}(\epsilon)$ for a time $t \sim \Delta^{-1}$ with a cost that is $\mathcal{O}(\text{poly}(\Delta^{-1}, N, \log(1/\epsilon)))$.

For $K = C \log(N)$, implementing the controlled time evolution is slightly more difficult. The Hamiltonian is no longer sparse when expressed in the computational basis. However, we can still use the approach of Ref. 23 or the related approach in Ref. 28 using a sequence of measurements to implement time evolution as follows. The term $-sB(X/N)^K$ can be expressed as a sum of terms, each of which is proportional to a projector onto an eigenspace of X, with eigenvalues $-N, -N + 2, \ldots, N$ so that there are $N + 1$ such terms. We can measure each such term using polynomially many gates by applying a Hadamard to all qubits, use an adder to compute to total X, then uncomputing the addition and undoing the Hadamard. A similar approach might also work for the Taylor series approach[7]. Even without this way of decomposing $-sB(X/N)^K$, the overhead due to the non-sparsity of the Hamiltonian would be quasi-polynomial and would be negligible compared to the improvements in theorems 1,2.

We choose the adiabatic evolution so that for input state $\psi = \psi_{0,1}$, the output state has squared overlap with $\psi_{0,0}$ equal to $1 - P_{diab}$ for some P_{diab} close to 0. From the adiabatic theorems quoted above, the adiabatic evolution time required is only polynomial in P_{diab}^{-1}. A similar estimate comes from using the repeated measurements of Ref. 28. Hence,

Lemma 1. *For any ψ, and $P_{diab} > 0, \epsilon > 0$, the algorithm succeeds with probability at least $|\langle\psi_{0,1}|\psi\rangle|^2(1-P_{diab})-\mathcal{O}(\epsilon)$. It takes a time $\mathcal{O}(\text{poly}(N, \log(1/\epsilon), P_{diab}^{-1}))$. Assuming success, the output state has squared overlap at least $1 - \epsilon$ with $\psi_{0,0}$.*

Choosing P_{diab} and ϵ small we obtain the desired P_{succ} close to 1.

In fact, for the Hamiltonian H_s from section II, we will show later that (under a spectral gap assumption and under some assumptions on B, K that we give later that)

$$|\langle\psi_{0,1}|\psi_{0,0}\rangle|^2 = \Omega(1). \tag{9}$$

Hence, for this problem, we can use a single measurement, phase estimating $H_{0,1}$, and not use any adiabatic evolution. There is no need for the final phase estimation on H_0 since this measurement in the computational basis projects into an eigenstate of H_Z whose energy can then be computed classically.

This leads to the following simple algorithm Algorithm 3 which replaces Algorithm 1.

As a technical remark, the above algorithm assumes that we know $E_{0,1}$. However, in the application that we consider, we do have good estimates on $E_{0,1}$, given later.

As a further remark, for even D many of the Hamiltonian simulation algorithms that we refer to can be chosen to preserve the even subspace exactly. Even if they do not preserve it exactly, they preserve it up to some error ϵ which is chosen to be negligible. Hence, if desired, one can do the simulation using H_s *not* projected into the even subspace.

Algorithm 3 Simplified Short-Path (unamplified version)

---

1. Let $\psi = \psi_+$ be the input state.

2. Phase estimate ψ using Hamiltonian H_1. If the energy estimate is greater than $E_{0,s} + \Delta/2$, then terminate the algorithm and return failure.

3. Measure the state in the computational basis and compute the value of H_Z after measuring. If this value is equal to E_0 then declare success and output the computational basis state.

IV. BRILLOUIN-WIGNER PERTURBATION THEORY

We use Brillouin-Wigner perturbation theory to compute the ground state of H_s as a function of s. In the case that H_0 has a unique ground state (as we consider here), Brillouin-Wigner perturbation theory gives a particularly simple result for the ground state of H_s, given in Eq. (13). There are many references for Brillouin-Wigner perturbation theory; a useful reference is Ref. 21, which also gives some convergence results. We will however derive the needed results below.

For even D, we work in the even subspace throughout this section. We remark on the appropriate choice of basis states later.

A. Introduction

We begin with some general results on Brillouin-Wigner perturbation theory. The results in this subsection do not use any properties of the specific choice of H_s above, except that we assume that H_0 has a unique ground state (one great feature of this perturbation theory, however, is that it becomes only slightly more complicated when the ground state is degenerate, while Rayleigh-Schrodinger perturbation theory becomes much more complicated). We consider a Hamiltonian $H_s = H_0 + sV$ in this subsection with H_0, V arbitrary, later taking $V = -B(X/N)^K$.

Let $|0\rangle$ be the ground state of H_0. Let $Q = 1 - |0\rangle\langle 0|$. Let $\phi_{0,s}$ be the ground state of H_s. To define the Brillouin-Wigner perturbation theory, it is convenient to normalize the states differently, rather than normalizing them to have unit norm. We will use the normalization that

$$\langle \phi_{0,s}|0\rangle = 1. \tag{10}$$

so that

$$\psi_{0,s} = \frac{\phi_{0,s}}{|\phi_{0,s}|}. \tag{11}$$

Let $E_{0,s}$ denote the energy of $\psi_{0,s}$ for Hamiltonian H_s. Let

$$G_0(\omega) = (Q(\omega - H_0)Q)^{-1}, \tag{12}$$

where ω is a scalar and where the inverse is computed in the subspace which is the range of Q and let $(1 - Q)G_0 = G_0(1 - Q) = 0$. That is, $G_0(\omega)$ is a Moore-Penrose pseudo-inverse of $Q(\omega - H_0))Q$, so that $G_0(\omega)(\omega - H_0) = (\omega - H_0)G_0(\omega) = Q$.

Then Brillouin-Wigner perturbation theory gives

$$\phi_{0,s} = |0\rangle + \sum_{k \geq 1} \Big(sG_0(E_{0,s})V \Big)^k |0\rangle. \tag{13}$$

Note that by definition $G_0(\omega)V = G_0(\omega)QV$.

Note that $E_{0,s}$ appears in the power series of $\phi_{0,s}$. Thus, in applications of this perturbation theory to compute eigenvectors or eigenvalues, it is necessary to self-consistently compute $E_{0,s}$, using

$$E_{0,s} = E_0 + s \sum_{k \geq 0} \langle 0|V \Big(sG_0(E_{0,s})V \Big)^k |0\rangle \tag{14}$$

We will give bounds on $E_{0,s}$ later.

The correctness of Eq. (13) as a formal power series in s can be readily verified by computing $(H - E_{0,s})\phi_{0,s}$. In subsequent subsections, we will give conditions for convergence of this power series and we will compute $|\phi_{0,s}|$. Eq. (10) is immediate from Eq. (13).

B. Overlap

Now consider the specific choices of $H_0 = H_Z$ and $V = -B(X/N)^K$ using Eq. (6) We will compute the overlap $\langle \psi_+ | \phi_{0,1} \rangle$, assuming convergence of the series (13) and assuming a bound on $E_0 - E_{0,1}$. Note that $E_{0,1} < E_0$.

For odd D, let us use $|u\rangle, |v\rangle, \ldots$ to denote basis states in the computational basis. For even D, we use $|u\rangle$ to denote the equal amplitude superposition of a pair of basis states in the computational basis, with the two basis states related by flipping all the spins. Each u labels a bit string of length N; let $\bar{u}$ denote the bit string with all bits flipped. Then, $|u\rangle$ and $|\bar{u}\rangle$ denote the same basis state for even D.

Let $E_u, E_v, \ldots$ denote the corresponding eigenvalues for these states for Hamiltonian H_0. Then,

$$
\begin{aligned}
\langle \psi_+ | \phi_{0,1} \rangle = {} & \langle \psi_+ | 0 \rangle \\
& + B \sum_{u \neq 0} \langle \psi_+ | u \rangle \frac{\langle u | (X/N)^K | 0 \rangle}{E_u - E_{0,1}} \\
& + B^2 \sum_{u \neq 0} \sum_{v \neq 0} \langle \psi_+ | u \rangle \frac{\langle u | (X/N)^K | v \rangle}{E_u - E_{0,1}} \frac{\langle v | (X/N)^K | 0 \rangle}{E_v - E_{0,1}} \\
& + \ldots
\end{aligned}
\tag{15}
$$

For any u, we have $\langle \psi_+ | u \rangle = 2^{-N/2}$.

Before continuing, we need two technical lemmas. First, the following lemma which estimates $\langle 0 | (X/N)^L | 0 \rangle$.

Lemma 2. *For $0 < L < N$ even, $\langle 0 | (X/N)^L | 0 \rangle \leq L!!/N^L \leq (L/N)^{L/2}$ where $(L-1)!! = (L-1) \cdot (L-3) \cdot \ldots$.*
For L odd, $\langle 0 | (X/N)^L | 0 \rangle = 0$.
For L, L' both even with $0 < L < L'$ we have $\langle 0 | (X/N)^L | 0 \rangle > \langle 0 | (X/N)^{L'} | 0 \rangle$, hence for $L > N/2$ we have $\langle 0 | (X/N)^L | 0 \rangle \leq 2^{-N/4}$.

Proof. We have $\langle 0 | (X/N)^L | 0 \rangle = N^{-L} \sum_{i_1, \ldots, i_L} \langle 0 | X_{i_1} \ldots X_{i_L} | 0 \rangle$.

First consider odd D. The expectation value vanishes unless for all j, there are an even number of a such that $i_a = j$. In that case, the expectation value is equal to 1. Thus, there must be some $a > 1$ such that $i_a = i_1$. There are N possible choices of i_1 and $L - 1$ possible choices of $a > 1$. Hence, $\langle 0 | (X/N)^L | 0 \rangle \leq ((L-1)/N) \langle (X/N)^{L-2} \rangle$. So, $\langle 0 | (X/N)^L | 0 \rangle \leq L!!/N^L \leq (L/N)^{L/2}$, where $L!! = (L-1)(L-3) \ldots$.

For even D, for $L \geq N$, there are additional terms in the expectation where for all j, there are an *odd* number of a such that $i_a = j$. However, since we have chosen $L < N$, such terms do not occur.

The monotonic decrease with increasing L is immediate when working in the X basis since $(X/N)^L > (X/N)^{L'}$. $\square$

Second we need the following inequality

Lemma 3. *Let $x_1, \ldots, x_k$ be positive random variables. The variables need not be independent of each other. Then $\mathbb{E}[\prod_{i=1}^k x_i^{-1}] \geq \prod_{i=1}^k (\mathbb{E}[x_i])^{-1}$.*

Proof. We have $\mathbb{E}[\prod_{i=1}^k x_i^{-1}] = \mathbb{E}[\exp(-\sum_{i=1}^k \ln(x_i))] \geq \exp(-\mathbb{E}[\sum_{i=1}^k \ln(x_i)])$, where the inequality is by convexity of the exponential. However, $-\mathbb{E}[\ln(x_i)] \geq -\ln(\mathbb{E}[x_i])$, again by convexity. So, $\mathbb{E}[\prod_{i=1}^k x_i^{-1}] \geq \exp(-\sum_{i=1}^k \ln(\mathbb{E}[x_i])) = \prod_{i=1}^k (\mathbb{E}[x_i])^{-1}$. $\square$

We now show that

Lemma 4. *Assume that $E_{0,1} \geq E_0 - 1$ and assume that series (15) is convergent. Let $B = \mathcal{O}(\mathrm{poly}(N))$. Then,*

$$
\langle \psi_+ | \phi_{0,1} \rangle \geq 2^{-N/2} \left(\exp\left[\frac{BN}{(2DK + \mathcal{O}(1/N^3)) |E_0|} \right] - o(1) \right) \frac{1}{\mathrm{poly}(N)}.
\tag{16}
$$

Proof. Note that all terms in Eq. (15) are non-negative. We re-express the series in terms of a random walk on the basis states $|u\rangle$ as follows. The random walk starts in state $|0\rangle$ at time 0. If the random walk is in some state $|u_t\rangle$ at time t, then the state of the random walk at time $t+1$ is given by repeating K times the process of picking a random spin and flipping that spin. Note that we can flip the same spin more than once in a single step of the random walk (indeed, it may be flipped up to K times) although this is unlikely for $K << \sqrt{N}$. That is, each step of the random walk we consider here is K steps of a random walk on the Boolean hypercube.

For $t > 0$ we say that the random walk "returns at time t" if the state of the random walk at time t is $|0\rangle$. For $t > 0$ we say that the random walk "returns by time t" if the random walk returns at some time s with $0 < s \leq t$.

Let $P_{nr}(t)$ denote the probability that the random walk does not return by time t. Let $\mathbb{E}_{nr,t}$ denote an expectation value conditioned on the random walk not returning by time t. Then, we have

$$\langle\psi_+|\phi_{0,1}\rangle = 2^{-N/2}\sum_{t=0}^{\infty}B^t\mathbb{E}_{nr,t}\Big[\prod_{m=1}^{t}\frac{1}{E_{u_m}-E_{0,1}}\Big]P_{nr,t}. \tag{17}$$

where the random walk has a sequence of states $u_1,\ldots,u_t$.

We can estimate $P_{nr,t}$ from lemma 2. We have

$$P_{nr,t} \leq \sum_{0<s\leq t}\langle 0|(X/N)^{Ks}|0\rangle,$$

where $\langle 0|(X/N)^{Ks}|0\rangle$ is the probability that it returns at time s. By lemma 2, for $t = \mathcal{O}(\mathrm{poly}(N))$ and $K \geq 3$, we have $P_{nr,t} = \mathcal{O}(1/N^3)$.

Eq. (17) requires computing the expectation value of $\prod_{m=1}^{t}\frac{1}{E_{u_m}-E_{0,1}}$. Applying lemma 3 to this expectation value we have

$$\mathbb{E}_{nr,t}\Big[\prod_{m=1}^{t}\frac{1}{E_{u_m}-E_{0,1}}\Big] \geq \prod_{m=1}^{t}\frac{1}{\mathbb{E}_{nr,t}[E_{u_m}-E_{0,1}]}. \tag{18}$$

Suppose the state of the random walk at time t is given and has some energy E_{u_t}. Then, if we pick a single spin at random and flip it, the expectation value of the energy of the resulting state is equal to

$$\Big(1-2\frac{D}{N}\Big)E_{u_t}.$$

To see this, consider any term in H_0 which is a product of D spins; the probability that we flip one of these spins, changing the sign of this term, is D/N. Remark: this is the point at which we use that all terms in H_Z have the same degree D; otherwise, the dependence of the average energy on m may be more complicated.

Repeating mK times, we find that the expectation value of the energy at time $m+1$, for given u_m, is equal to

$$\mathbb{E}[E_{u_m}] = \Big(1-2\frac{D}{N}\Big)^{mK}E_0 \tag{19}$$
$$\leq \Big(1-2\frac{DmK}{N}\Big)E_0.$$

Since $E_{u_m} \geq E_0$ for all u_m, we have for $t = \mathcal{O}(\mathrm{poly}(N))$,

$$\mathbb{E}[E_{u_m}] \geq (1-P_{nr,t})E_0 + P_{nr,t}\mathbb{E}_{nr,t}[E_{u_m}]. \tag{20}$$

So, for $t = \mathcal{O}(\mathrm{poly}(N))$ where $P_{nr,t} = \mathcal{O}(1/N^3)$ we have

$$\mathbb{E}_{nr,t}[E_{u_m}] \leq \Big(1-2\frac{DmK}{N}-\mathcal{O}(1/N^3)\Big)E_0. \tag{21}$$

So, by Eqs. (18,21) we have

$$\mathbb{E}_{nr,t}\Big[\prod_{m=1}^{t}\frac{1}{E_{u_m}-E_{0,1}}\Big] \tag{22}$$

$$\geq \frac{1}{1+(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}\cdot\frac{1}{1+(\frac{4DK}{N}+\mathcal{O}(1/N^3))|E_0|}\cdots\frac{1}{1+(\frac{2tDK}{N}+\mathcal{O}(1/N^3))|E_0|}$$

$$\geq \Big(\frac{1}{(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}\Big)^t\cdot\frac{1}{t!}\cdot\frac{1}{1+\frac{1}{(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}}\cdot\frac{1}{1+\frac{2}{(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}}\cdots\frac{1}{1+\frac{t}{(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}}$$

$$\geq \Big(\frac{1}{(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}\Big)^t\cdot\frac{1}{t!}\cdot\exp\Big(-\frac{\log(t)+1}{1+\frac{1}{(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}}\Big).$$

We may assume that $|E_0| = \Omega(N)$ as otherwise the ground state is degenerate. So, for t only polynomially large, the quantity $\exp\left(-\frac{\log(t)+1}{1+\frac{1}{(\frac{2DK}{N}+\mathcal{O}(1/N^3))|E_0|}}\right)$ is only polynomially small; indeed for $|E_0| = \omega(N\log(N))$, this quantity is $\Omega(1)$. So, the sum in Eq. (15) obeys

$$\langle \psi_+ | \phi_{0,1} \rangle \geq 2^{-N/2} \sum_{t=0}^{\mathcal{O}(\text{poly}(N))} B^t P_{nr,t}\left(\frac{N}{(2DK+\mathcal{O}(1/N^3))|E_0|}\right)^t \frac{1}{t!} \frac{1}{\text{poly}(N)} \tag{23}$$

$$= 2^{-N/2}\left(\exp\left[\frac{BN}{(2DK+\mathcal{O}(1/N^3))|E_0|}\right] - o(1)\right)\frac{1}{\text{poly}(N)}.$$

Here, we have used the fact that the power series expansion of $\exp(\alpha)$ is given by $\sum_{t\geq 0} \alpha^t/t!$ and the fact that $\sum_{t\geq s} \alpha^t/t!$ is negligible for $s >> \alpha$. Indeed, $\sum_{t\geq s} \alpha^t/t!$ is exponentially small in α for fixed ratio s/α with $s/\alpha > e$. Hence, we can choose the polynomial in the limits on the first line of Eq. (23) to be large compared to $\frac{BN}{(2DK+\mathcal{O}(1/N^3))|E_0|}$ and then the sum of remaining terms (i.e., the terms in the series expansion of the exponential which are *not* included in the first line of Eq. (23)) is $o(1)$. $\qquad\square$

V. CONVERGENCE PROPERTIES, ENERGY SHIFT, AND NORM

In this subsection, we consider convergence of the series (15), we bound the shift on energy $E_0 - E_{0,1}$ and we bound the norm $|\phi_{0,1}|$, and we consider the gap of H_s. Most of the results will be based on considering the spectrum of QH_sQ. Again we take $H_0 = H_Z$ and $V = -B(X/N)^K$.

We first show the following three results. Eq. (27) below implies that Eq. (9) holds.

Lemma 5. *1. The series (15) always converges, assuming that the value of $E_{0,1}$ in the series indeed is equal to the ground state energy of H_1.*

2. Consider the Hamiltonian QH_sQ. Let $E_{0,s}^Q$ be the smallest eigenvalue of this Hamiltonian in the subspace spanned by the range of Q. Then, $\partial_s E_{0,s}^Q \leq 0$.

3. Finally, assume that $E_{0,1}^Q \geq E_0 + 1/2$. Then,

$$E_{0,1} \geq E_0 - B\langle 0|(X/N)^K|0\rangle - 2B^2\langle 0|(X/N)^{2K}|0\rangle. \tag{24}$$

and

$$|\phi_{0,1}|^2 \leq 1 + 4B^2\left|Q(X/N)^K|0\rangle\right|^2 \tag{25}$$

$$\leq 1 + 4B^2\langle 0|(X/N)^{2K}|0\rangle.$$

For K odd with $\langle 0|B^2(X/N)^{2K}|0\rangle \leq 1/2$, we have that

$$E_{0,1} \geq E_0 - 1, \tag{26}$$

and

$$|\phi_{0,1}| \leq 2. \tag{27}$$

Proof. Defining

$$G_s(\omega) = \left(Q(\omega - H_s)Q\right)^{-1}, \tag{28}$$

where again the inverse is a Moore-Penrose inverse, the power series (13) is a series expansion of

$$\phi_{0,s} = |0\rangle + sG_s(E_{0,s})V|0\rangle \tag{29}$$

in powers of s. We now consider the singularities of $G_s(\omega)$ considered as a function of s. Then, the radius of convergence of the series (13) is equal to the distance from the origin $s = 0$ to the closest singularity (as a function of s) of $G_s(\omega)$ in the complex plane.

First, we will show that the singularities are simple poles at values of s where QH_sQ has an eigenvector in the range of Q with eigenvalue equal to ω. Let $g = (Q(\omega - H_0)Q)^{-1/2}$ and let $v = QVQ$, where the exponent $-1/2$ means a square-root of the Moore-Penrose inverse. For ω smaller than all eigenvalues of H_0 (which holds for our problem since $E_{0,1} < E_0$), this inverse is well-defined and the square-root can be taken pure imaginary so that g is anti-Hermitian. We choose the sign of the square-root arbitrarily. Then,

$$G_s(\omega) = g(1 - sgvg)^{-1}g. \tag{30}$$

To see Eq. (30), we have $Q(\omega - H_0 - sV)Qg(1 - sgvg)^{-1}g = Q(g^{-1} - sVg)(1 - sgvg)^{-1}g = Qg^{-1}(1 - sgVg)(1 - sgvg)^{-1}g = Q$, where all inverses are Moore-Penrose inverses. The matrix gvg is Hermitian and so can be diagonalized. Thus, the singularities of $G_s(\omega)$ are simple poles at values of s equal to eigenvalues of $(gvg)^{-1}$. Note that while elsewhere in the paper we assume that $s \in [0,1]$, here we allow s to be an arbitrary complex number. Let ζ be an eigenvector of gvg with eigenvalue s^{-1}, so that $(1 - sgvg)\zeta = 0$. Hence, $0 = g(g^{-1} - svg)\zeta = g(g^{-2} - sv)g\zeta = g(\omega - H_s)g\zeta$ so that $g\zeta$ is an eigenvector of QH_sQ with eigenvalue ω, as claimed.

The above results about singularities hold for arbitrary H_0, V, ω. Now to show convergence, we consider $H_0 = H_Z, V = -B(X/N)^K, \omega = E_{0,1} \leq E_0$. Suppose some matrix element of $G_s(E_{0,1})$ in the computational basis has radius of convergence r. Every matrix element of $G_s(E_{0,1})$ in the computational basis has a series expansion with all coefficients the same sign. So, by Pringsheim's theorem, there must be a singularity for $s = r$, i.e., there must be a singularity on the positive real axis. So, the series for $G_1(E_{0,1})$ converges unless $E_{0,t}^Q = E_{0,1}$ for some $t \in [0,1]$.

Since all off-diagonal terms of QH_sQ and QVQ are negative, we have $\partial_s E_{0,s}^Q \leq 0$ by Perron-Frobenius, thus proving item 2. Hence, the series for $s = 1$ is convergent if $E_{0,1}^Q > E_{0,1}$. However, $E_{0,1}^Q > E_{0,1}$ also by Perron-Frobenius, thus proving item 1.

With this definition of G_s, we have

$$E_{0,s} = E_0 + s\langle 0|V|0\rangle + s^2\langle 0|VG_s(E_{0,s})V|0\rangle. \tag{31}$$

Let $H_0 = H_Z, V = -B(X/N)^K$. Assume that $E_{0,1}^Q \geq E_0 + 1/2$. Since $E_{0,1} \leq E_0$, $G_1(E_{0,1})$ has operator norm bounded by 2. So,

$$E_{0,1} \geq E_0 - B\langle 0|(X/N)^K|0\rangle - 2B^2\langle 0|(X/N)^{2K}|0\rangle. \tag{32}$$

Also, if $G_1(E_{0,1})$ has operator norm bounded by 2, we have

$$|\phi_{0,1}|^2 \leq 1 + 4B^2\Big|Q(X/N)^K|0\rangle\Big|^2 \tag{33}$$
$$\leq 1 + 4B^2\langle 0|(X/N)^{2K}|0\rangle.$$

This proves Eqs. (24,25). Eqs. (26,27) are immediate. $\qquad\square$

We now consider the gap of H_s:

Lemma 6. *The Hamiltonian H_s has a gap between ground and first excited states that is greater than or equal to $E_{0,s}^Q - E_0$. Since $\partial_s E_{0,s}^Q \leq 0$, the gap of H_s is greater than or equal to $E_{0,1}^Q - E_0$.*

Proof. This is a special case of a general result. Consider a Hamiltonian

$$H = E_0|0\rangle\langle 0| + \sum_{a \neq 0} E_a|a\rangle\langle a| + \sum_{a \neq 0} v_a\Big(|a\rangle\langle 0| + \text{h.c.}\Big), \tag{34}$$

where v_a is an arbitrary vector. We will show that for any E_a, v_a, E_0, the spectral gap of this Hamiltonian is greater than or equal to $\min_{a \neq 0}(E_a - E_0)$. Then, to apply this result to the spectral gap of H_s, we take the states $|a\rangle$ in Eq. (34) to be eigenstates of QH_sQ and take v_a to be matrix elements of sV between $|0\rangle$ and those eigenstates.

Define the Green's function $G(\omega) = (\omega - H)^{-1}$. We have

$$\langle 0|G|0\rangle = \Big(\omega - E_0 - \Sigma(\omega)\Big)^{-1}, \tag{35}$$

where

$$\Sigma(\omega) = \sum_{a \neq 0} |v_a|^2(\omega - E_a)^{-1}. \tag{36}$$

For $\omega < \min_{a \neq 0} E_a$, we have $\Sigma(\omega) < 0$. Hence, $\langle 0|G|0 \rangle$ does not have any poles in the interval $E_0 < \omega < \min_{a \neq 0} E_a$. Hence, if H has an eigenvalue in this interval, then the corresponding eigenvector has vanishing amplitude on $|0\rangle$; however, any such eigenvector has eigenvalue equal to E_a for some a, so no such eigenvector exists.

Thus, all eigenvalues of H, for any v are contained in $(-\infty, E_0] \cup [\min_a(E_a), \infty)$. If $v_a = 0$ for all a, there is exactly one eigenvalue in the interval $(-\infty, E_0]$ and so this cannot change as v_a changes. $\qquad\square$

Hence it follows that:

Theorem 3. *Consider the Hamiltonian QH_sQ. Let $E_{0,s}^Q$ be the smallest eigenvalue of this Hamiltonian in the subspace spanned by the range of Q. Assume that $E_{0,1}^Q \geq E_0 + 1/2$ and assume that $\langle 0|B^2(X/N)^{2K}|0\rangle \leq 1/2$. Then*

1. *The Hamiltonian H_s has gap at least $1/2$ between the ground and first excited state.*

2. *We have*

$$\langle \psi_+ | \psi_{0,1} \rangle \; \geq \; \frac{1}{2} 2^{-N/2} \Big(\exp\Big[\frac{BN}{(2DK + \mathcal{O}(1/N^3))|E_0|} \Big] - o(1) \Big) \frac{1}{\text{poly}(N)}. \tag{37}$$

Remark: for $B = -bE_0$, $J_{tot} = \mathcal{O}(N^\beta)$, $K \geq \beta$, the condition $\langle 0|B^2(X/N)^{2K}|0\rangle \leq 1/2$ holds for all sufficiently large N.

VI. GAP ASSUMPTION AND ENTROPY

In this section we assume that $E_{0,1}^Q < E_0 + 1/2$ and prove some consequences of that. In subsection VI A we show how to construct states with large expectation value for X. In subsection VI B, we show entropic consequences of this using a log-Sobolev inequality. In subsection VI C, we put these results together.

For technical convenience later, it is easier to work with an eigenvector of H_1. So we show:

Lemma 7. *Assume $E_{0,1}^Q < E_0 + 1/2$. Then, there is an eigenvector Ψ of H_1 with eigenvalue at most $E_0 + 1/2$ such that $\langle \Psi | B(X/N)^K | \Psi \rangle \geq 1/4$.*

Proof. H_1 has at least two eigenvalues with energy at most $E_0 + 1/2$. This is because the ground state of QH_1Q in the subspace spanned by Q and the state $|0\rangle$ span a two dimensional space. Further, the average of energy over the corresponding eigenvectors is at most $E_0 + 1/4$ since the ground state energy of H_1 is $\leq E_0$. On the other hand, the average of H_Z over these eigenvectors is at least $E_0 + 1/2$. Hence, the average of $B(X/N)^K$ over these eigenvectors is at least $1/4$. So, at least one such Ψ exists. $\qquad\square$

Remark: the above lemma is the only place in this section in which the degeneracy of H_Z is used. We used the degeneracy to show that the average of H_Z over these eigenvectors is at least $E_0 + 1/2$. However, it is not hard to remove the degeneracy assumption here, at the cost of slightly worse constants. We sketch this in the discussion.

A. Localizing in X

In this section, we show how given an eigenvector with a large expectation value for $B(X/N)^K$ we can construct a wavefunction with a large expectation value for X and whose expectation value for H_Z is only slightly changed. This will be necessary to apply log-Sobolev bounds in the next section.

We begin with a lemma that considers states of a Hamiltonian describing a single particle hopping in one-dimension:

Lemma 8. *Let h be a real Hermitian tridiagonal matrix. Label rows and columns of h by an integer x, so that $h_{x,y} = 0$ if $|x - y| > 1$. Let ψ be a real eigenvector of h with eigenvalue E. Let $\psi(x)$ denote the x-th entry of ψ. Let h_{od} denote the off-diagonal part of h, i.e, h_{od} has the same entries as h off the main diagonal and is zero on the main diagonal. Then,*

1. *For any integer $\ell > 0$, for any y, there is a state ξ with $|\xi| = 1$ such that*

$$\langle \xi | h | \xi \rangle \leq E + \mathcal{O}(1/\ell^2) \| h_{od} \|, \tag{38}$$

and such that $\xi(x)$ is non-vanishing only for $x < y + \ell$ or $\xi(x)$ is non-vanishing for $x > y - \ell$.

2. *For small $\epsilon > 0$, the state ψ_ϵ defined by $\psi_\epsilon(x) = \exp(\epsilon x)\psi(x)$ obeys*

$$\frac{\langle \psi_\epsilon | h | \psi_\epsilon \rangle}{|\psi_\epsilon|^2} \leq E + \mathcal{O}(\epsilon^2)\|h_{od}\|. \tag{39}$$

3. *If the state ψ_ϵ defined above has*

$$\sum_{x \geq z} \psi_\epsilon(x)^2 \geq \frac{1}{2}\sum_x \psi_\epsilon(x)^2 \tag{40}$$

for some z, then there is a state ξ with $|\xi| = 1$ such that

$$\langle \xi | h | \xi \rangle \leq E + \mathcal{O}(1/\ell^2 + \epsilon^2)\|h_{od}\| \tag{41}$$

and such that either $\xi(x)$ is non-vanishing only for $y - \ell < x < y + \ell$ for some $y \geq z$.

Proof. Assume without loss of generality that $E = 0$ (otherwise add a scalar to h so that $E = 0$). Assume without loss of generality that $\psi(x)$ is real. Let $f(x)$ be any function. Let $\hat{f}$ be the diagonal matrix with entries given by $f(\cdot)$. We compute

$$\langle \hat{f}\psi | h | \hat{f}\psi \rangle \tag{42}$$
$$= \sum_x \Big(f(x)^2 h_{x,x}\psi(x)^2 + f(x)f(x+1)h_{x,x+1}\psi(x)\psi(x+1) + f(x)f(x-1)h_{x,x-1}\psi(x)\psi(x-1) \Big)$$
$$= \sum_x f(x)^2 \psi(x)\Big(h_{x,x}\psi(x) + h_{x,x-1}\psi(x-1) + h_{x,x+1}\psi(x+1) \Big)$$
$$\quad + \sum_x \Big[f(x)(f(x+1) - f(x))h_{x,x+1}\psi(x)\psi(x+1) + f(x)(f(x-1) - f(x))h_{x,x-1}\psi(x)\psi(x-1) \Big]$$
$$= \sum_x \Big[f(x)(f(x+1) - f(x))h_{x,x+1}\psi(x)\psi(x+1) + f(x)(f(x-1) - f(x))h_{x,x-1}\psi(x)\psi(x-1) \Big]$$
$$= \sum_x \Big[f(x)(f(x+1) - f(x))h_{x,x+1}\psi(x)\psi(x+1) + f(x+1)(f(x) - f(x+1))h_{x+1,x}\psi(x+1)\psi(x) \Big]$$
$$= -\sum_x (f(x) - f(x+1))^2 h_{x,x+1}\psi(x)\psi(x+1),$$

where the third equality follows from the fact that $h_{x,x}\psi(x) + h_{x,x-1}\psi(x-1) + h_{x,x+1}\psi(x+1) = 0$ since ψ is an eigenvector with eigenvalue 0 and the fourth equality follows by shifting the variable in the summation by 1.

To prove the first claim, let $\rho^< = \sum_{y-\ell < x \leq y}\psi(x)^2$ and let $\rho^> = \sum_{y \leq x < y+\ell}\psi(x)^2$. Assume $\rho^< > \rho^>$. Choosing $f(x) = \ell$ for $x \leq y$ and $f(x) = \ell - (x-y)$ for $y < x < y+\ell$ and $f(x) = 0$ for $x \geq y+\ell$, we find that $\langle \hat{f}\psi | h | \hat{f}\psi \rangle \leq \|h_{od}\|\rho^>$ and $|\hat{f}\psi|^2 \geq \ell^2\rho^<$, so that for $\xi = \hat{f}\psi/|\hat{f}\psi|$, Eq. (38) is satisfied and $\xi(x)$ is nonvanishing only for $x < y+\ell$. If instead $\rho^< \leq \rho^>$, choose instead $f(x) = \ell$ for $x \geq y$ and $f(x) = \ell - |x - y|$ for $y - \ell < x < y$ and $f(x) = 0$ for $x \leq y - \ell$ and for $\xi = \hat{f}\psi/|\hat{f}\psi|$, Eq. (38) is satisfied and $\xi(x)$ is nonvanishing only for $x > y - \ell$.

To prove the second claim, choose $f(x) = \exp(\epsilon x)$, so that we have $\langle \hat{f}\psi || h | \hat{f}\psi \rangle = -\sum_x \psi(x)\psi(x+1)h_{x,x+1}\mathcal{O}(\epsilon^2)\exp(2\epsilon x) \leq \mathcal{O}(\epsilon^2)\|h_{od}\| \cdot |\hat{f}\psi|^2$.

To prove the third claim, define $\rho_y = \sum_{|x-y| \leq \ell/2} \exp(2\epsilon x)\psi(x)^2$. If the assumption (40) holds, then there must be some $y > z$ such that $\rho_y \geq (1/2)\rho_{y-\ell}$; to see this, suppose no such y exists, then $\rho_{z+\ell/2} < (1/2)\rho(z - \ell/2)$ and $\rho(z + 3\ell/2) < (1/4)\rho(z + \ell/2)$ and so on, so $\sum_{x \geq z}\psi_\epsilon(x)^2 < \sum_{x < z}\psi_\epsilon(x)^2$.

So, there must be some $y > z$ such that $\rho_y \geq (1/2)\rho_{y-\ell}$ and such that $\rho_y \geq (1/2)\rho_{y+\ell}$. To see this, find the largest $y > z$ such that $\rho_y \geq (1/2)\rho_{y-\ell}$; by the above paragraph at least one such y must exist. This y must have $\rho_y \geq (1/2)\rho_{y+\ell}$ as otherwise it would not be the largest.

Choose $f(x) = \exp(\epsilon x)\ell$ for $|y - x| \leq \ell$, choose $f(x) = \exp(\epsilon x)(2\ell - |y - x|)$ for $\ell \leq |y - x| \leq 2\ell$ and choose $f(x) = 0$ otherwise. $\qquad\square$

The above lemma applies to a one-dimensional Hamiltonian. However, we can apply it to an eigenvector Ψ of the Hamiltonian H_1 as follows. Let P_x project onto the eigenspace of X with eigenvalues in the interval $[xD, xD + D)$. That is, defining $q(y) = \lfloor y/D \rfloor$, then P_x projects onto the eigenspace of X with eigenvalues y such that $q(y) = x$.

Then, the Hamiltonian H_1, projected into the space spanned by $P_x\Psi/|P_x\Psi|$ obeys the conditions of lemma 8 with $\|h_{od}\| \leq \|H_Z\|$ and with $\psi(x) = \sqrt{\langle P_x\Psi|\Psi\rangle}$.

We now show how to attain a wavefunction with a large expectation value for X. We do this in two slightly different ways depending on how large K is. The lemma will involve the constant $C > 0$; the constants hidden inside the big-O notation do *not* depend upon C.

Lemma 9. *Assume that $E_{0,1}^Q < E_0 + 1/2$.*

1. *Assume that $K = C\log(N)$. Let $B = -bE_0$ with $b \leq 1$. Then, there is a state Ξ, with $|\Xi| = 1$ such that*

$$\langle\Xi|X/N|\Xi\rangle \geq 1 - \mathcal{O}(1)/C \tag{43}$$

and such that

$$\langle\Xi|H_Z|\Xi\rangle \leq E_0 + 1/2 + \mathcal{O}(1) \cdot \frac{J_{tot}}{N^2}C^2D^2\log(N)^2 + B. \tag{44}$$

2. *Consider arbitrary K. Let $B = -bE_0$ with $b \leq 1$. Let $X_{min} = N \cdot (4B)^{-1/K}$. (Remark: for $X < X_{min}$ we have $B(X/N)^K \leq 1/4$.)*

Then, there is a state Ξ with $|\Xi| = 1$ such that Ξ is supported on an eigenspace of X with eigenvalues in some interval $[X_0 - X_{min}/K, X_0 + X_{min}/K]$ for $X_0 \geq X_{min}$ and such that

$$\langle\Xi|H_Z - B(X/N)^K|\Xi\rangle \leq E_0 + 1/2 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2}, \tag{45}$$

so that

$$\begin{aligned}\langle\Xi|H_Z|\Xi\rangle &\leq E_0 + 1/2 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B\Big(\frac{X_0 + X_{min}/K}{N}\Big)^K \\ &\leq E_0 + 1/2 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + eB(X_0/N)^K.\end{aligned} \tag{46}$$

Proof. The proof of both cases is the same. By lemma 7, there is an eigenstate Ψ of H_1 with $\langle\Psi|H_1|\Psi\rangle \leq E_0 + 1/2$ and $\langle\Psi|B(X/N)^K|\Psi\rangle \geq 1/4$. Throughout, the state ψ will be an eigenvector of h constructed from Ψ as explained above by projecting into the space spanned by $P_x\Psi/|P_x\Psi|$.

Let $X_{min} = N \cdot (4B)^{-1/K}$. For $X < X_{min}$ we have $B(X/N)^K < 1/4$.

Recall that K is chosen odd. So $\exp(K(X - X_{min})/X_{min}) \geq 4B(X/N)^K$, and so

$$\exp(KX/X_{min}) \geq 4\exp(K)B(X/N)^K. \tag{47}$$

Hence, since $\langle\Psi|B(X/N)^K|\Psi\rangle \geq 1/4$, we have

$$\langle\Psi|\exp(KX/X_{min})|\Psi\rangle \geq \exp(K). \tag{48}$$

Recall that $q(y) = \lfloor y/D\rfloor$. Construct the state ψ_ϵ of lemma 8 using $\epsilon = KD/2X_{min}$. Then,

$$|\psi_\epsilon|^2 \geq \langle\Psi|\exp(KDq(X)/X_{min})|\Psi\rangle \geq \exp(-2\epsilon)\langle\Psi|\exp(KX/X_{min})|\Psi\rangle \geq \exp(-2\epsilon)\exp(K),$$

using Eq. (48). At the same time, $\sum_{x<q(X_{min}-\ln(2)X_{min}/K)}|\psi_\epsilon(x)|^2 \leq (1/2)\exp(K - 2\epsilon)$. So, the state ψ_ϵ obeys Eq. (40) for $z = q(X_{min} - \ln(2)X_{min}/K)$. Then construct state ξ using item **3** of lemma 8 choosing $\ell = \text{const.} \times \epsilon^{-1}$, using $\|h_{od}\| \leq J_{tot}$.

From state ξ, construct state $\Xi = \sum_x \xi(x)P_x\Psi/|P_x\Psi|$. We have chosen $\ell = \text{const.} \times \epsilon^{-1}$; by choosing this constant smaller than $(1 - \ln(2))$, Ξ will be supported on the given eigenspace of X. This shows Eq. (45). Eq. (44) follows because in this case $X_{min} = N \cdot (1 - \mathcal{O}(1)/C)$. $\square$

B.   Log-Sobolev Inequality

Given a quantum state ψ, let $S^{(comp)}(\psi)$ be the entropy of the probability distribution of measurement outcomes when measuring the state in the computational basis. That is, if $\psi = \sum_u \psi(u)|u\rangle$, with $\sum_u |\psi(u)|^2 = 1$, then

$$S^{(comp)}(\psi) = -\sum_u |\psi(u)|^2 \log(|\psi(u)|^2). \tag{49}$$

Here we measure entropy using bits, i.e., taking logs to base 2, rather than nats. This will lead to some various $\ln(2)$ differences between our definitions and definitions in the log-Sobolev literature.

In this subsection, we relate the entropy $S^{(comp)}(\psi)$ to $\langle\psi|X|\psi\rangle$. It is clear that if $\langle\psi|X|\psi\rangle = N$, then $\psi = \psi_+$ up to an overall phase, and so $S^{(comp)}(\psi) = N$. Roughly, we will show that if $\langle\psi|X|\psi\rangle$ is extensive (i.e., equal to N times some nonzero constant), then $S^{(comp)}(\psi)$ is also extensive.

One such result uses the log-Sobolev inequality. This lemma 10 is not tight: it only gives a nontrivial bound on $S^{(comp)}(\psi)$ if $\langle\psi|X|\psi\rangle > (1 - \ln(2))N$. In lemma 11 we give a tight bound and give a precise statement of that rough extensivity claim above.

Lemma 10. *Let $\psi(u)$ be real. Then,*

$$S^{(comp)}(\psi) \geq \left(1 - \frac{1}{\ln(2)}\right) N + \frac{1}{\ln(2)} \langle\psi|X|\psi\rangle. \tag{50}$$

Proof. This follows from the log-Sobolev inequality[13,24]. Define for any function $f(u)$,

$$\mathrm{Ent}(f) = \mathbb{E}[f \log(\frac{f}{\mathbb{E}[f]})], \tag{51}$$

where the expectation is taken for a random choice of u in the domain of f. We have $\mathbb{E}[\psi^2] = 2^{-N} \sum_u \psi(u)^2 = 2^{-N}$. Here we abuse notation to use ψ to represent both a quantum state and a function $\psi(u)$. So,

$$\begin{aligned} \mathrm{Ent}(\psi^2) &= 2^{-N} \sum_u \psi(u)^2 \Big(N + \log(\psi(u)^2)\Big) \\ &= 2^{-N}\Big(N - S^{(comp)}(\psi)\Big). \end{aligned} \tag{52}$$

The log-Sobolev inequality states that

$$\mathrm{Ent}(\psi^2) \leq 2^{-N} \frac{N - \langle\psi|X|\psi\rangle}{\ln(2)}. \tag{53}$$

Hence Eq. (50) follows.

Remark: the log-Sobolev literature usually expresses the inequality in terms of gradients of a function on the Boolean hypercube; however, since this gradient is defined in terms of bit flips and X induces a single bit flip, after a little algebra one may see that Eq. (53) is equivalent to the results in the literature. $\qquad\square$

A tighter bound on $S^{(comp)}$ follows from the log-Sobolev inequality of Ref. 30. We have

Lemma 11. *Let $S(x) = -x\log(x) - (1-x)\log(1-x)$ be the binary entropy function (we use S rather than H to avoid confusion with the Hamiltonian H). Let*

$$\tau(\sigma) = 2\sqrt{S^{-1}(\sigma)\Big(1 - S^{-1}(\sigma)\Big)}. \tag{54}$$

(The inverse of S may be chosen arbitrarily so long as the same inverse is chosen in both locations.) Then,

$$\tau\Big(\frac{S^{(comp)}(\psi)}{N}\Big) \geq \frac{\langle\psi|X|\psi\rangle}{N}. \tag{55}$$

The function $\tau(\cdot)$ is a continuous increasing function, taking $[0, 1]$ to $[0, 1]$. For small σ,

$$\tau(\sigma) = \Theta(\sqrt{\frac{\sigma}{-\log(\sigma)}}), \tag{56}$$

and for small $\langle\psi|X|\psi\rangle/N$,

$$\frac{S^{(comp)}(\psi)}{N} = \Theta((X/N)^2 \log(N/X)). \tag{57}$$

Proof. This follows from theorem 1.2 of Ref. 30. That theorem is expressed in terms of $\text{Ent}(\psi^2)$, up to multiplicative factors of $\ln(2)$ since natural logs are used in that paper. Use Eq. (52) to re-express this in terms of $S^{(comp)}(\psi)$ and then Eq. (55) follows after some algebraic manipulations. □

The function $\tau(\cdot)$ has a simple interpretation. Consider the single qubit state

$$\sqrt{S^{-1}(\sigma)}|0\rangle + \sqrt{1 - S^{-1}(\sigma)}|1\rangle.$$

A product state of N qubits, all in this single qubit state, has entropy $N\sigma$ and expectation value of X equal to $2N\sqrt{S^{-1}(\sigma)(1 - S^{-1}(\sigma))}$. The result is that this product state minimizes entropy for the given expectation value of X.

Remark: in fact, while we have only considered the case of pure states, it is possible to show a generalization of Eq. (55) to mixed states. We do not need this generalization here, but it is interesting to note that it exists. Let ρ be a density matrix. Let $S^{(comp)}(\rho)$ be the entropy of the mixed state obtained by measuring ρ in the computational basis. Then, we have

$$\tau\Big(\frac{S^{(comp)}(\rho)}{N}\Big) \geq \frac{\text{tr}(X\rho)}{N}. \tag{58}$$

This can be proven similarly to the proof of the pure state case: one establishes it for a single qubit, and then one uses convexity of c and conditional entropy to show it for an arbitrary number of qubits.

C. Number of Low Energy Eigenstates

Theorem 4. *Assume that* $E_{0,1}^Q < E_0 + 1/2$.

1. *Assume that* $K = C\log(N)$. *Let* $B = -bE_0$ *with* $b \leq 1$. *Let* $C = \Theta(1)$. *Then, there is some probability distribution* $p(u)$ *on computational basis states with entropy at least*

$$S^{(comp)} \geq N \cdot (1 - \mathcal{O}(1)/C)$$

and with expected value of H_Z *at most* $(1 - b)E_0 + \mathcal{O}(1) \cdot \frac{J_{tot}}{N^2} C^2 D^2 \log(N)^2$.

2. *Consider arbitrary* K. *Let* $B = -bE_0$ *with* $b \leq 1$. *Then, for some* $X_0 \geq X_{min} = N \cdot (4B)^{-1/K}$, *there is some probability distribution* $p(u)$ *on computational basis states with entropy at least*

$$S^{(comp)} \geq N\tau^{-1}((X_0 - X_{min}/K)/N)$$

and with expected value of H_Z *at most*

$$E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B\Big(\frac{X_0 + X_{min}/K}{N}\Big)^K \cdot \mathcal{O}(1),$$

where the function $\tau(\cdot)$ *is defined in lemma 11. If* $X_0 << N$, *then*

$$2^{N\tau^{-1}(X_0/N)} = 2^{\Omega((X_0/N)^2 \log(N/X_0))N}$$

.

Proof. This follows from lemmas 9,10,11. We have dropped the additive $+1/2$ from lemma 9 as it is smaller than other terms hidden in the big-O notation. □

VII.   PROOF OF THEOREMS 1,2

We now prove the main theorem. Either $E_{0,1}^Q \geq E_0 + 1/2$, in which case theorem 3 applies or $E_{0,1}^Q < E_0 + 1/2$ in which case theorem 4 applies. As noted, for $B = -bE_0$, for $K \geq 3$ and $K > \beta$, the condition $B(X/N)^K \leq 1/2$ in theorem 3 holds for all sufficiently large N.

Theorem 4 shows the existence of a probability distribution with large entropy and small expectation value for H_Z. To turn this into a statement about $W(E)$ as in theorems 1,2 we need the following lemma. Remark: the factor of $\mathcal{O}(\log(N))$ in the statement of the lemma can be interpreted, for physicists, as an entropy difference arising when passing from a canonical to a microcanonical ensemble, and the need for two energies E_1, E_2 can be interpreted as a Maxwell construction.

Lemma 12. *Let $p(u)$ be a probability distribution over basis states $|u\rangle$ such that*

$$\sum_u p(u)E_u = \overline{E}. \tag{59}$$

Then, there are two energies, E_1, E_2 with $E_1 \leq \overline{E} \leq E_2$ such that for some probability $P \in [0,1]$ we have $PE_1 + (1-P)E_2 = \overline{E}$ and $P\log(W(E_1)) + (1-P)\log(W(E_2)) \geq S - \mathcal{O}(\log(N))$, where $W(E)$ is the number of computational basis states with expectation value E for H_Z.

Proof. Let

$$P(E) = \sum_{u:E_u=E} p(u). \tag{60}$$

If $P(E) > 0$, then for u such that $E_u = E$, let $p(u|E) = p(u)/P(E)$ and let

$$S(E) = - \sum_{u:E_u=E} p(u|E)\log(p(u|E)). \tag{61}$$

Then,

$$S = - \sum_E P(E)\log(P(E)) + \sum_E P(E)S(E). \tag{62}$$

Since there are only $O(J_{tot})$ possible values of E for which $P(E) \neq 0$, we have $-\sum_E P(E)\log(P(E)) = \mathcal{O}(\log(N))$. Hence, $\sum_E P(E)S(E) \geq S - \mathcal{O}(\log(N))$. Also, $\sum_E P(E)E = \overline{E}$. We have $S(E) \leq \log(W(E))$. Hence, $\sum_E P(E)\log(W(E)) \geq S - \mathcal{O}(\log(N))$.

Now maximize $\sum_E P(E)\log(W(E))$ subject to the linear constraints $\sum_E P(E)E = \overline{E}$ and $\sum_E P(E) = 1$, with $0 \leq P(E) \leq 1$ for all E. Introducing Lagrange multipliers λ_1, λ_2 corresponding to these constraints, one finds that for all E we have either $P(E) = 0$ or $P(E) = 1$ (and hence only one choice of E has $P(E) > 0$) or $\log(W(E)) = \lambda_1 E + \lambda_2$. Let T be the set of E such that $\log(W(E)) = \lambda_1 E + \lambda_2$. Since this constraint is linear, for any choice of $P(E)$ such that $P(E) = 0$ for $E \notin T$ and $\sum_E P(E) = 1$ and $\sum_P(E)E = \overline{E}$ we have the same $\sum_E P(E)\log(W(E))$. Hence, to maximize $\sum_E P(E)\log(W(E))$ it suffices to consider the case that $P(E)$ is nonvanishing for at most 2 choices of E. Let these choices be E_1, E_2 with $E_1 < E_2$ and $E_1 \leq \overline{E} \leq E_2$. $\square$

Remark: In the lemma above we used the fact that H_Z has integer eigenvalues to bound the entropy $-\sum_E P(E)\log(P(E))$. For more general H_Z, where J_{ij} are chosen more generally, we could bin energies into polynomially many bins and obtain a similar result.

Using this lemma, now consider the first case of theorem 4, where $K = C\log(N)$. Theorem 4 shows the existence of a probability distribution $p(u)$ on computational basis states with entropy at least $S = S^{(comp)} \geq N \cdot (1 - \mathcal{O}(1)/C)$ and with expected value of H_Z at most $\overline{E} = (1-b)E_0 + 1/2 + \mathcal{O}(1) \cdot \frac{J_{tot}}{N^2}C^2D^2\log(N)^2$. By lemma 12, there are two energies, E_1, E_2 with $E_1 \leq \overline{E} \leq E_2$ such that for some probability $P \in [0,1]$ we have $PE_1 + (1-P)E_2 = \overline{E}$ and $P\log(W(E_1)) + (1-P)\log(W(E_2)) = S - \mathcal{O}(\log(N))$. If $\log(W(E_1)) \geq S - \mathcal{O}(\log(N))$, the conclusion follows. Otherwise, we must have $\log(W(E_2)) \geq S - \mathcal{O}(\log(N))$. If $E_2 - E_0 \geq (1+\eta)(\overline{E} - E_0)$, then $P_2 \leq 1/(1+\eta)$ since $E_1 - E_0 \geq 0$. Since $\log(W(E_2)) \leq N$, we have $P_1\log(W(E_1)) + P_2N \geq S - \mathcal{O}(\log(N))$ so $\log(W(E_1)) \geq N \cdot (1 - \mathcal{O}(1)\frac{1+\eta}{\eta}\frac{1}{C}) - \frac{1+\eta}{\eta}\mathcal{O}(\log(N))$.

This completes the proof of theorem 1.

Now consider the second case of theorem 4, where K is a constant. The theorem shows that for some $X_0 \geq X_{min} = N \cdot (4B)^{-1/K}$, there is some probability distribution $p(u)$ on computational basis states with entropy at

least $S = S^{(comp)} \geq N\tau^{-1}((X_0 - X_{min}/K)/N)$ and with expected value of H_Z at most $\overline{E} = E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B(\frac{X_0+X_{min}/K}{N})^K \cdot \mathcal{O}(1)$. Shifting X_0 by X_{min}/K, we find that for some $X_0 \geq (1 - 1/K)N \cdot (4B)^{-1/K}$, there is some probability distribution $p(u)$ on computational basis states with entropy at least $S = S^{(comp)} \geq N\tau^{-1}(X_0/N)$ and with expected value of H_Z at most $\overline{E} = E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B(\frac{X_0+2X_{min}/K}{N})^K \cdot \mathcal{O}(1)$. However, for the given range of X_0, we have $(\frac{X_0+2X_{min}/K}{N})^K \leq (X_0/N)^K\mathcal{O}(1)$, so $\overline{E} \leq E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B(X_0/N)^K \cdot \mathcal{O}(1)$. Now removing the restriction on the range of X_0 allowing $0 \leq X_0 \leq N$ and defining $m_x = X_0/N$ we find that for some $m_x \in [0,1]$ there is a probability distribution with entropy $S \geq N\tau^{-1}(m_x)$ and with expected value of H_Z at most

$$E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B(m_x)^K \cdot \mathcal{O}(1) \tag{63}$$
$$\leq E_0 + \mathcal{O}(1)\frac{J_{tot}K^2D^2}{X_{min}^2} + B(\tau(S/N))^K \cdot \mathcal{O}(1)$$
$$\equiv F(S).$$

By lemma 12, there are two energies, E_1, E_2 with $E_1 \leq \overline{E} \leq E_2$ such that for some probability $P \in [0,1]$ we have $PE_1 + (1-P)E_2 = \overline{E} \leq F(S)$ and $P\log(W(E_1)) + (1-P)\log(W(E_2)) = S - \mathcal{O}(\log(N))$. The function $F(S)$ is a convex function, so either $E_1 \leq F(\log(W(E_1)) + \mathcal{O}(\log(N))))$ or $E_2 \leq F(\log(W(E_2)) + \mathcal{O}(\log(N))))$. This completes the proof of theorem 2.

VIII.   DISCUSSION

We have given a quantum algorithm for exact optimization. This algorithm leads to a nontrivial speedup, assuming a bound on the number of computational basis states with low energy for H_Z. This naturally leads to a hybrid algorithm: if the bound is obeyed, we run the exact algorithm here while if the bound is not obeyed, one can find a low energy state (i.e., an approximate solution of the optimization problem) by repeated random sampling or by Grover search.

We have only considered problems where all terms in H_Z have the same degree. Eq. (19) is the point at which we used this assumption. For problems such as MAX-3-SAT, the energy dependence may be more complicated so it is not clear how well this algorithm works there. It would be interesting to study this using numerical simulation. Also we emphasize that the mere presence of a large number of low energy states of H_Z is not, on its own, sufficient to make the algorithm fail; rather, one needs to have a low energy eigenstate of QH_1Q, which requires some structure to how those states are arranged on the hypercube. That is, if a large number of low energy states are distributed in some fairly random way on the hypercube it is possible that one might not have a small $E_{0,1}^Q$. Numerical simulation may also be useful here.

Lemma 7 is the only place in section VI where we used the assumption on the degeneracy of H_Z. However, it is not hard to remove this assumption here. Assume $E_{0,1}^Q < E_0 + 1/2$. Then consider the Hamiltonian $H_Z - (5/2)B(X/N)^K$. This must have an eigenvalue $\leq E_0 - 1/4$. Hence there must be some eigenvector Ψ of $H_Z - (5/2)B(X/N)^K$ with $\langle\Psi|B(X/N)^K|\Psi\rangle \geq 1/4$. Then one could use this Ψ in the rest of section VI; the cost is that in some cases factors of B in this section get replaced by $(5/2)B$, in particular in theorem 4. One can equivalently understand this as follows: one instead assumes that $E_{0,2/5}^Q < E_0 + 1/2$. If this holds, then $E_{0,1}^Q < E_0 - 1/4$. If this assumption fails, then section 3 implies a nontrivial speedup for the algorithm using Hamiltonian $H_Z - (2/5)B(X/N)^K$. In fact, there is nothing special about our assumption that $E_{0,1}^Q < E_0 + 1/2$. The analysis of section V would work under the assumption that $E_{0,1}^Q \geq E_0 + c$ for any constant $c > 0$. Further, there is nothing special about our assumption in the lemma that $\langle\Psi|B(X/N)^K|\Psi\rangle \geq 1/4$; the analysis above would go through up to some changes in constants if we assumed that $\langle\Psi|B(X/N)^K|\Psi\rangle \geq c'$ for any $c' > 0$. So, it would suffice to consider $H_Z - c''B(X/N)^K$ for $c'' = (1+c')/(1-c)$ which can be made arbitrarily close to 1. In future work, we will address the degeneracy assumption in section V.

---

[1] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 307–316. ACM, 2015. doi:10.1137/15m1050902.

[2] Boris Altshuler, Hari Krovi, and Jérémie Roland. Anderson localization makes adiabatic quantum optimization fail. *Proceedings of the National Academy of Sciences*, 107(28):12446–12450, 2010. doi:10.1073/pnas.1002116107.

[3] Andris Ambainis and Oded Regev. An elementary proof of the quantum adiabatic theorem. *arXiv preprint quant-ph/0411152*, 2004.

[4] Charles H Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM journal on Computing*, 26(5):1510–1523, 1997. doi:10.1137/s0097539796300933.

[5] D. W. Berry, A. M. Childs, and R. Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 792–809, Oct 2015. doi:10.1109/FOCS.2015.54.

[6] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Exponential improvement in precision for simulating sparse Hamiltonians. pages 283–292, 2014. doi:10.1145/2591796.2591854.

[7] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Simulating Hamiltonian dynamics with a truncated Taylor series. *Phys. Rev. Lett.*, 114:090502, 2015. doi:10.1103/PhysRevLett.114.090502.

[8] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002. doi:10.1090/conm/305.

[9] A Crisanti and T Rizzo. Analysis of the ∞-replica symmetry breaking solution of the sherrington-kirkpatrick model. *Physical Review E*, 65(4):046137, 2002. doi:10.1103/physreve.65.046137.

[10] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516):472–475, 2001. doi:10.1126/science.1057726.

[11] Martin Fürer and Shiva Prasad Kasiviswanathan. Exact max 2-sat: Easier and faster. In *International Conference on Current Trends in Theory and Practice of Computer Science*, pages 272–283. Springer, 2007. doi:10.1007/978-3-540-69507-3_22.

[12] Alexander Golovnev and Konstantin Kutzkov. New exact algorithms for the 2-constraint satisfaction problem. *Theoretical Computer Science*, 526:18–27, 2014. doi:10.1016/j.tcs.2014.01.010.

[13] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. doi:10.2307/2373688.

[14] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219. ACM, 1996. doi:10.1145/237814.237866.

[15] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical review letters*, 103(15):150502, 2009. doi:10.1103/physrevlett.103.150502.

[16] M. B. Hastings. Local maxima and improved exact algorithm for max-2-sat. *Chicago Journal of Theoretical Computer Science*, 2018. doi:10.4086/cjtcs.2018.002. arXiv preprint arXiv:1610.07100.

[17] S Kirkpatrick and D Sherrington. Solvable model of a spin-glass. *Phys. Rev. Lett*, 35(26):1792–1796, 1975. doi:10.1103/physrevlett.35.1792.

[18] A Yu Kitaev. Quantum measurements and the abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995.

[19] Alexei Yu Kitaev, Alexander Shen, and Mikhail N Vyalyi. *Classical and quantum computation*, volume 47. American Mathematical Society Providence, 2002. doi:10.1090/gsm/047.

[20] Sergey Knysh and Vadim Smelyanskiy. On the relevance of avoided crossings away from quantum critical point to the complexity of quantum adiabatic algorithm. *arXiv preprint arXiv:1005.3011*, 2010.

[21] JM Leinaas and TTS Kuo. Convergence properties of the brillouin-wigner type of perturbation expansion. *Annals of Physics*, 111(1):19–37, 1978. doi:10.1016/0003-4916(78)90222-1.

[22] Guang Hao Low and Isaac L Chuang. Hamiltonian simulation by qubitization. 2016.

[23] Guang Hao Low and Isaac L. Chuang. Optimal Hamiltonian simulation by quantum signal processing. *Phys. Rev. Lett.*, 118:010501, 2017. doi:10.1103/PhysRevLett.118.010501.

[24] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014. doi:10.1017/cbo9781139814782.002.

[25] Sergey Pankov. Low-temperature solution of the sherrington-kirkpatrick model. *Physical review letters*, 96(19):197204, 2006. doi:10.1103/physrevlett.96.197204.

[26] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979. doi:10.1103/physrevlett.43.1754.

[27] Giorgio Parisi. Order parameter for spin-glasses. *Physical Review Letters*, 50(24):1946, 1983. doi:10.1103/physrevlett.50.1946.

[28] D. Poulin, A Kitaev, D. S. Steiger, M. B. Hastings, and M. Troyer. Fast quantum algorithm for spectral properties. *arXiv preprint:arXiv:1711.11025*, 2017.

[29] Ben W Reichardt. The quantum adiabatic optimization algorithm and local minima. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 502–510. ACM, 2004. doi:10.1145/1007352.1007428.

[30] Alex Samorodnitsky. A modified logarithmic sobolev inequality for the hamming cube and some applications. *arXiv preprint arXiv:0807.1679*, 2008.

[31] Alexander D Scott and Gregory B Sorkin. Linear-programming design and analysis of fast algorithms for max 2-csp. *Discrete Optimization*, 4(3-4):260–287, 2007. doi:10.1016/j.disopt.2007.08.001.

[32] RD Somma, S Boixo, Howard Barnum, and E Knill. Quantum simulations of classical annealing processes. *Physical review letters*, 101(13):130504, 2008. doi:10.1103/physrevlett.101.130504.

[33] Dave Wecker, Matthew B Hastings, and Matthias Troyer. Training a quantum optimizer. *Physical Review A*, 94(2):022309, 2016. doi:10.1103/physreva.94.022309.

[34] Ryan Williams. A new algorithm for optimal 2-constraint satisfaction and its implications. *Theoretical Computer Science*, 348(2-3):357–365, 2005. doi:10.1016/j.tcs.2005.09.023.

Appendix A: Reduction to Unique Ground State

We now corollaries how to reduce the problem without the degeneracy assumption to the problem with the degeneracy assumption, at an additional time cost.

For given H_Z, consider a more general family of Hamiltonians, of the form $H_Z + \sum_i h_i Z_i$, where the h_i are chosen from $\{-1, 0, +1\}$. We will define a sequence of such Hamiltonians, $H^{(0)}, H^{(1)}, \ldots, H^{(m)}$. We will set $H^{(0)} = H_Z$ so that for $H^{(0)}$ all $h_i = 0$. Then, each $H^{(a+1)}$ is constructed from $H^{(a)}$ by picking some i such that $h_i = 0$ in the Hamiltonian $H^{(a)}$ and setting h_i either equal to $+1$ or -1 in $H^{(a+1)}$, while leaving all other terms in the Hamiltonian unchanged.

Before continuing we need the following lemma:

Lemma 13. *Consider any arbitrary set S with $S \subseteq \{-1, +1\}^N$ and $|S| > 1$. Write elements of S as vectors $v = (v_1, \ldots, v_N)$ with $v_i \in \{-1, +1\}$. Then, there exists some $i \in \{1, \ldots, N\}$ and some $\sigma \in \{-1, +1\}$ such that for*

$$T = S \cap \{v | v_i = \sigma\}$$

we have $1 \leq |T| \leq S/2$. In words, the set T is the set of vectors in S such that the i-th entry of the vector is equal to σ.

Proof. For each i, let $n_i = |S \cap \{v | v_i = +1\}|$. If for some i we have $0 < n_i < |S|$, then either $1 \leq n_i \leq |S|/2$ in which case we can pick $T = S \cap \{v | v_i = +1\}$ or $S/2 \leq n_i \leq |S| - 1$ in which case we can pick $T = S \cap \{v | v_i = -1\}$. On the other hand, if for all i we have $n_i \in \{0, |S|\}$, then $|S| \leq 1$ which contradicts the assumptions of the lemma (proof: without loss of generality assume that $n_i = |S|$ for all i; then, only the all $+1$ vector can be in S). $\square$

Let $H^{(a)}$ have $n_{gs}(a)$ different ground states. Write $n_{gs} = n_{gs}(0)$. We now show

Lemma 14. *For some $m \leq \log(n_{gs})$, we can choose a sequence $H^{(0)}, \ldots, H^{(m)}$ such that $1 \leq n_{gs}(a + 1) \leq n_{gs}(a)/2$ for all $0 \leq i \leq a - 1$ with $n_{gs}(m) = 1$.*

Proof. The proof is inductive. Let $S^{(a)}$ be the set of ground states of $H^{(a)}$. We will construct each $H^{(a)}$ such that if it has $h_i \neq 0$ for any i, then all ground states of $H^{(a)}$ have $v_i = -h_i$. Then, apply lemma 13 with $S = S^{(a)}$. Find the i, σ so that for $T = S \cap \{v | v_i = \sigma\}$ we have $1 \leq |T| \leq S/2$. By the inductive assumption $h_i = 0$ in Hamiltonian $H^{(a)}$ as if $h_i \neq 0$ then $T = \emptyset$ or $T = S^{(a)}$. Then set $H^{(a+1)} = H^{(a)} - \sigma Z_i$. Since the added term $-\sigma Z_i$ is equal to its minimal possible value of -1 on some non-empty subset of the ground states of $H^{(a)}$, those states are the ground states of $H^{(a+1)}$ and the inductive assumption holds. Continue in this fashion, until for some a we have $n_{gs}(a) = 1$. $\square$

Hence, if H_Z has n_{gs} ground state, then there exists some Hamiltonian $H_Z + \sum_i h_i Z_i$ with the following properties. First, it has a unique ground state which is also a ground state of H_Z. Second, the number of nonzero h_i is at most $\log(n_{gs})$. That is, writing $\vec{h}$ as a vector with entries h_i, the vector $\vec{h}$ has $\log(n_{gs})$ nonzero entries.

There are

$$2^1 \binom{N}{1} + 2^2 \binom{N}{2} + \ldots + 2^{n_{gs}} \binom{N}{\log(n_{gs})} \leq (2N)^{\log(n_{gs})}$$

possible choices of $\vec{h}$.

The Hamiltonians of the form $H_Z + \sum_i h_i Z_i$ do not have all terms of the same degree, as some terms have degree D and some have degree 1. However, for any such Hamiltonian $H_Z + \sum_i h_i Z_i$ with a unique ground state we can define a Hamiltonian H' on $N + D + 1$ qubits which obeys the degeneracy assumption (i.e., has a unique ground state for odd D or a doubly degenerate state for even D) so that the ground state of $H_Z + \sum_i h_i Z_i$ can be trivially obtained from a ground state of H'. To do this, add $D + 1$ qubits, labelled $N + 1, \ldots, N + D + 2$. Write J equal to the sum of all possible D-th order monomials in Pauli Z operators on those added qubits, with a coefficient -1 in front of each monomial. For example, for $D = 3$, we have $J = -Z_{N+1}Z_{N+2}Z_{N+3} - Z_{N+1}Z_{N+2}Z_{N+4} - Z_{N+1}Z_{N+3}Z_{N+4} - Z_{N+2}Z_{N+3}Z_{N+4}$. Consider the Hamiltonian

$$H_Z + J + \sum_{i \leq N} h_i Z_i Z_{N+1} Z_{N+2} \ldots Z_{N+D-1}.$$

The term J is minimized for even D, by all Z_i being the same for $i > N$, while for odd D is it minimized by all Z_i equaling $+1$ for $i > N$. The terms $H_Z + \sum_{i \leq N} h_i Z_i Z_{N+1} Z_{N+2} \ldots Z_{N+D-1}$ can be minimized by choosing all $Z_i = +1$ and choosing the unique ground state of $H_Z + \sum_i h_i Z_i$ found above. Thus $H_Z + J + \sum_{i \leq N} h_i Z_i Z_{N+1} Z_{N+2} \ldots Z_{N+D-1}$ obeys the degeneracy assumption and finding its ground state directly gives a ground state of H_Z.

Thus, we can reduce to the unique ground state problem at the cost of trying $(2N)^{\log(n_{gs})}$ different choices of $\vec{h}$. This number is small if n_{gs} is small, while if n_{gs} is large, one can find the ground state more rapidly by Grover search. Considering the algorithm with logarithmically growing K, note that if $n_{gs} \geq 2^{rN/\log(N)^2}$ for some scalar r then one can find the ground state using Grover search in expected time $\mathcal{O}^*(2^{N/2-rN/\log(N)^2/2})$, while if $n_{gs} \leq 2^{rN/\log(N)^2}$ then there are only $(2N)^{rN/\log(N)^2} = 2^{rN/\log(N)+\mathcal{O}(N/\log(N)^2)}$ choices of $\vec{h}$. For $r < b/(2CD)$, theorem 1 shows a speedup greater than $2^{rN/\log(N)}$ and so one still has a $2^{\text{const.}\times N/\log(N)}$ speedup in this case, so in worst case one has a $2^{\text{const.}\times N/\log(N)^2}$ speedup.