Sensitivity of quantum speedup by quantum annealing to a noisy oracle

Siddharth Muthukrishnan, Tameem Albash, and Daniel A. Lidar

Phys. Rev. A **99**, 032324 — Published 15 March 2019

DOI: 10.1103/PhysRevA.99.032324

Sensitivity of quantum speedup by quantum annealing to a noisy oracle

Siddharth Muthukrishnan,[1,2,*] Tameem Albash,[1,2,3] and Daniel A. Lidar[1,2,4,5]

[1] *Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, USA*
[2] *Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*
[3] *Information Sciences Institute, University of Southern California, Marina Del Rey, California, USA*
[4] *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*
[5] *Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA*

The glued-trees problem is the only example known to date for which quantum annealing provides an exponential speedup, albeit by partly using excited state evolution, in an oracular setting. How robust is this speedup to noise on the oracle? To answer this, we construct phenomenological short-range and long-range noise models, and noise models that break or preserve the reflection symmetry of the spectrum. We show that under the long-range noise models an exponential quantum speedup is retained. However, we argue that a classical algorithm with an equivalent long-range noise model also exhibits an exponential speedup over the noiseless model. In the quantum setting the long-range noise is able to lift the spectral gap of the problem so that the evolution changes from diabatic to adiabatic. In the classical setting, long-range noise creates a significant probability of the walker landing directly on the EXIT vertex. Under short-range noise the exponential speedup is lost, but a polynomial quantum speedup is retained for sufficiently weak noise. In contrast to noise range, we find that breaking of spectral symmetry by the noise has no significant impact on the performance of the noisy algorithms. Our results about the long-range models highlight that care must be taken in selecting phenomenological noise models so as not to change the nature of the computational problem. We conclude from the short-range noise model results that the exponential speedup in the glued-trees problem is not robust to noise, but a polynomial quantum speedup is still possible.

I. INTRODUCTION

Quantum annealing (QA) [1–5] usually refers to a family of analog quantum optimization algorithms that interpolate between an initial Hamiltonian whose ground state is easy to prepare and a final Hamiltonian whose ground state is the answer to the optimization problem we want to solve [6]. Typically, QA is operated adiabatically, which means that the interpolation timescale t_f (also referred to as the annealing time) is much larger than the smallest energy gap between the ground state and the first excited state that is encountered along the interpolation. The adiabatic theorem for closed system dynamics provides a guarantee that for a sufficiently long t_f, the evolution reaches the ground state of the final Hamiltonian with high probability (see, e.g., Ref. [7] for a rigorous statement).

We can also consider QA operated non-adiabatically. Here too, the goal is to end the evolution with the system in the ground state of the final Hamiltonian, but the system can undergo diabatic transitions to excited states and return to the ground state. To further complicate matters, QA can also refer to a version of open system analog quantum optimization algorithms operating at non-zero temperature [8].

In this work, we consider a particular diabatic, oracular QA algorithm for solving the glued-trees problem, which we modify by the addition of noise to the oracle. The glued-trees problem was first introduced in Ref. [9], where it was shown that any classical algorithm must necessarily take exponential time to solve this problem, and a quantum walk algorithm was presented which solves the problem in polynomial time. Subsequently, a diabatic QA algorithm was presented which also solves the problem in polynomial time [10]. This is so far the only explicit QA algorithm for which an exponential speedup is known. The QA evolution in the algorithm takes the system from the ground state to the first excited state, then back down to the ground state. This transition from and back to the ground state is enabled by the Hamiltonian spectrum, which is symmetric about the middle of evolution.

Oracular models may not be practical examples of quantum speedups because it is highly non-trivial to construct an oracle in a way that does not assume that we already know the answer to the problem at hand; and, even if we could do so, oracular Hamiltonians acting on n spins typically involve n-body operators. However, they provide insights into the mechanisms and boundaries of quantum speedups, and can sometimes serve as stepping stones to more practical, non-oracular algorithms [11, 12]. In this work, we address the question of whether the exponential speedup of the QA glued-trees algorithm is robust under noise. The noise models we consider are phenomenological and add a time-independent random matrix with Gaussian entries to the interpolating Hamiltonian. Such noise is more appropriately viewed as a model of control errors than as originating from a system-bath interaction [13]. We consider two dichotomies of noise models. One dichotomy is between noise mod-

* muthukri@usc.edu

els which induce long-range interactions among distant nodes in the graph and noise models which only induce interactions between nearest-neighbor nodes. The other dichotomy is between noise models which break a certain reflection symmetry in the spectrum and noise models which preserve the reflection symmetry.

Our noise models are motivated by three concerns. First, they offer ways to perturb features of the problem that are considered explanatorily relevant to the performance of the QA algorithm. This will become clear later, but the main idea can be illustrated as follows. The QA algorithm described in Ref. [10] works reliably because the spectrum is symmetric upon reflection about the middle of the evolution. This symmetry guarantees that if the system is excited to the first excited state in the first half of the evolution due to the presence of an exponentially small energy gap, the system will then encounter the same exponentially small gap in the second half of the evolution and return back to the ground state. Therefore, a perturbation that breaks this reflection symmetry offers a control knob to explore the importance of this symmetry. Second, given that this is an oracle problem, in order to obtain physical noise models, we need to consider physical realizations of the oracle. But oracles are generically unrealizable as local Hamiltonians. Thus, in the absence of physical implementations, we assume the noise is Gaussian at the oracle level. Finally, we choose these noise models because they allow for a numerical and analytical treatment to reasonably large system sizes.

We now summarize our results. We find that for the long-range noise models, the quantum dynamics show an exponential speedup over classical algorithms that respect the glued-trees graph-structure. However, this speedup is misleading because an exponential speedup is also observed for a *classical* algorithm with long-range transition terms. More precisely, the long-range noise corresponds to a classical random walk on a graph containing edges connecting *any* two columns (see Fig. 1), which allows for a sufficiently high probability for the random walker to jump directly to the EXIT vertex. Meanwhile, we find that the quantum dynamics with the long-range noise exhibit a speedup because of a perturbative lifting of the spectral gap, which turns dynamics that were diabatic in the noiseless setting to dynamics that are adiabatic in the noisy setting. We also observe that the short-range noise models lose the exponential quantum speedup over the noiseless classical algorithm, but they do show a *polynomial* speedup for sufficiently small values of the noise strength [14].

The paper is organized as follows. In Sec. II, we describe the glued-trees problem and the QA algorithm that solves it. In Sec. III A, we describe the noise models that we study. In Sec. IV, we present numerical results on how the performance of the algorithm changes under the different noise models. In Sec. V, we provide an explanation for these results and we conclude in Sec. VI.

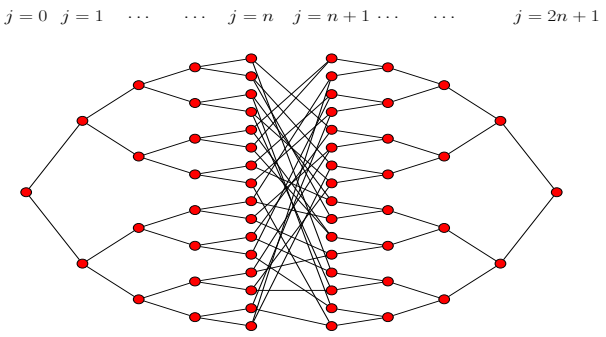


FIG. 1. (Color online) The graph structure of the glued trees problem. $j = 0, 1, 2, \ldots, n, n+1, \ldots, 2n+1$ indexes the different columns of the graph starting from one vertex to another.

II. THE GLUED-TREES PROBLEM

Consider two identical perfect binary trees, of depth n, glued together as depicted in Fig. 1. The gluing is done such that each leaf on one tree is randomly joined to two leaves on the other tree, and vice versa. This ensures that every vertex in the graph, except the two root vertices, have degree 3. One root node is called the ENTRANCE vertex and the other root node is called the EXIT vertex. Starting from the ENTRANCE vertex, the objective is to find the EXIT vertex [15].

All the vertices have labels. A classical algorithm can query the oracle with the label of any vertex, and the oracle returns the labels of the vertices connected to the given vertex. A quantum algorithm can query the oracle with the label of any vertex and the oracle will return a uniform superposition over all the vertices connected to it. Thus, the oracle encodes the adjacency matrix A of the glued-trees graph. Since n is the depth of one of the binary trees, the total number of vertices in the glued trees is $2^{n+2} - 2 = \mathcal{O}(2^n)$, which is the minimum number of distinct labels we need. Therefore, the entire graph can be labeled using $(n+2)$-length bitstrings. But this labeling system is insufficient to make the problem hard for classical algorithms. Rather, to prove classical hardness, there need to be exponentially more labels than vertices [9]. It turns out to be sufficient to choose labels as randomly chosen from the set of $2n$-length bitstrings.

Under this labeling scheme, Ref. [9] showed that any classical algorithm that makes fewer than $2^{n/6}$ queries to the oracle, will not be able to find the EXIT vertex with probability greater than $4 \times 2^{-n/6}$. This means that it will at least take a time $\Omega(2^{n/3})$ to find the EXIT vertex, because in order to boost the success probability we must repeat the algorithm $2^{n/6}$ times.

On the other hand, it was shown in Ref. [9] that a quantum walk algorithm which starts from the ENTRANCE vertex and evolves under the Hamiltonian equal to the adjacency matrix of the graph, can find the EXIT vertex

with probability $\mathcal{O}(\frac{1}{n})$ if the algorithm is run for times chosen uniformly at random in the interval $[0, \mathcal{O}(n^4)]$. This means we can can get a probability of success arbitrarily close to 1 by simply repeating the algorithm $\mathcal{O}(n)$ times, and therefore the algorithm will take at most $\mathcal{O}(n^5)$ time. This yields an exponential speedup over the classical algorithm [16].

A. The quantum annealing algorithm

We now turn to the QA algorithm for the glued-trees problem presented in Ref. [10]. The initial Hamiltonian is taken to be the projector onto the ENTRANCE vertex: $H_0 = -|\text{ENTRANCE}\rangle\langle\text{ENTRANCE}|$, such that the initial state of the system coincides with the ground state of the Hamiltonian. The final Hamiltonian is the projector onto the EXIT vertex: $H_1 = -|\text{EXIT}\rangle\langle\text{EXIT}|$. We then interpolate between these projectors while turning on and off the adjacency matrix A:

$$H(s) = (1-s)\alpha H_0 - s(1-s)A + s\alpha H_1, \tag{1}$$

with $s = t/t_f \in [0,1]$, where t is the physical time and t_f is the total evolution time. Also, $0 < \alpha < \frac{1}{2}$ is a constant. We set $\hbar = 1$ throughout. In Ref. [10], it was shown that if $t_f = \mathcal{O}(n^6)$, then the above interpolation ends with sufficiently high probability in the ground state of H_1, the EXIT vertex.

With the initial state being $|\text{ENTRANCE}\rangle$, the evolution associated with the Hamiltonian in Eq. (1) confines the system to the subspace spanned by the *column basis*, whose elements are defined as

$$|\text{col}_j\rangle \equiv \frac{1}{\sqrt{N_j}} \sum_{a \in \text{column } j} |a\rangle, \tag{2}$$

where $|a\rangle$ denotes the state associated with a vertex in column j with label a and

$$N_j = \begin{cases} 2^j, & 0 \le j \le n \\ 2^{2n+1-j}, & n+1 \le j \le 2n+1 \end{cases} \tag{3}$$

is the number of vertices in column j (there are $2n+2$ columns in total). It is straightforward to show (see Appendix A) that in the column basis, the matrix elements of the Hamiltonian [Eq. (1)] are

$$H_{0,0} = -\alpha(1-s) \tag{4a}$$
$$H_{j,j+1} = H_{j+1,j} = -s(1-s) \text{ for } j \neq n \tag{4b}$$
$$H_{n,n+1} = H_{n+1,n} = -\sqrt{2}s(1-s) \tag{4c}$$
$$H_{2n+1,2n+1} = -\alpha s. \tag{4d}$$

1. *Reflection symmetry*

This Hamiltonian is invariant under the composition of two transformations, which together we call the *reflection symmetry*. The first transformation is the reflection of the graph around the central glue. In the column basis, this is represented by the permutation matrix P which has 1's on the anti-diagonal and 0's everywhere else,

$$P_{ij} = \delta_{i,2n+1-j}, \quad i,j \in \{0,1,2,\ldots,(2n+1)\}. \tag{5}$$

The second transformation is $s \mapsto (1-s)$: the reflection of the interpolation parameter s around $s = 0.5$. The reflection symmetry is the invariance of the Hamiltonian [Eq. (1)] under the composition of these two transformations:

$$H(s) = PH(1-s)P. \tag{6}$$

One consequence of the reflection symmetry is that the spectrum of the Hamiltonian is symmetric under the second transformation $s \mapsto (1-s)$ alone. This is because Eq. (6) implies that $s \mapsto (1-s)$ corresponds to effectively conjugating the Hamiltonian by P, and since P is unitary, the spectrum is unchanged. Therefore,

$$E_k(s) = E_k(1-s) \text{ for } k \in \{0,1,2,\ldots,(2n+1)\}. \tag{7}$$

Another consequence of the symmetry is that if $|\phi_k(s)\rangle$ is the k-th eigenstate of $H(s)$, then

$$H(s)|\phi_k(s)\rangle = E_k(s)|\phi_k(s)\rangle \tag{8a}$$
$$\implies PH(s)P^\dagger P|\phi_k(s)\rangle = E_k(s)P|\phi_k(s)\rangle \tag{8b}$$
$$\implies H(1-s)(P|\phi_k(s)\rangle) = E_k(s)(P|\phi_k(s)\rangle) \tag{8c}$$
$$\implies H(1-s)(P|\phi_k(s)\rangle) = E_k(1-s)(P|\phi_k(s)\rangle) \tag{8d}$$
$$\implies |\phi_k(1-s)\rangle = P|\phi_k(s)\rangle. \tag{8e}$$

Together, Eqs. (7) and (8e) imply that $H(1-s)$ has the same eigenvalues as $H(s)$ and that the eigenvectors of $H(1-s)$ are the reversed-in-column-basis eigenvectors of $H(s)$.

B. Dynamics

As shown in Ref. [10], the key features of the Hamiltonian that results in a polynomial time performance are the scalings of the avoided level-crossings in the spectrum, depicted in Fig. 2. The evolution that solves the problem in polynomial time is as follows. The system starts in the ground state of the Hamiltonian at $s = 0$ (i.e., the ENTRANCE vertex). In the optimal evolution, the system diabatically transitions to the first excited state at the first exponentially small gap (between s_1 and s_2). Then, it adiabatically follows the first excited state and does not transition to the second excited state because of the polynomially large gap between the first and second excited states. Finally, the system returns back down to ground state through the second exponentially small gap (between s_3 and s_4). At the end of the annealing evolution described above, we get the EXIT vertex with high probability, as long as the evolution time t_f is chosen to scale as $\mathcal{O}(n^6)$.

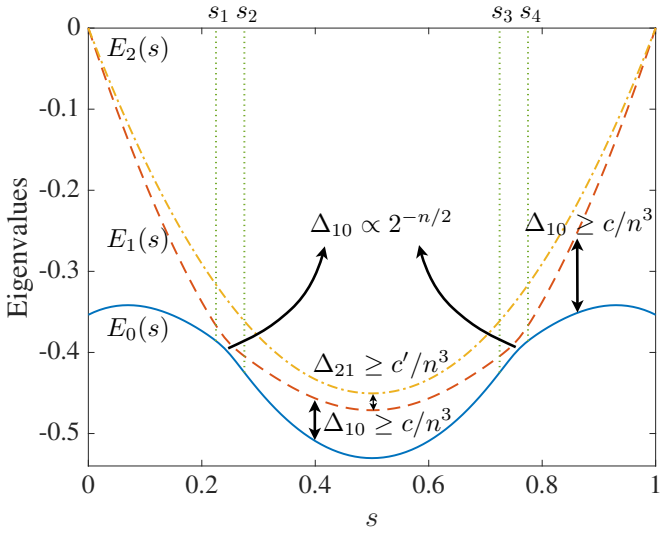


FIG. 2. (Color online) The smallest three eigenvalues of the Hamiltonian for the case of $n = 6$ and $\alpha = 1/\sqrt{8}$. We choose a small n so that the exponentially small gaps are visible. (This image from Ref. [6].)

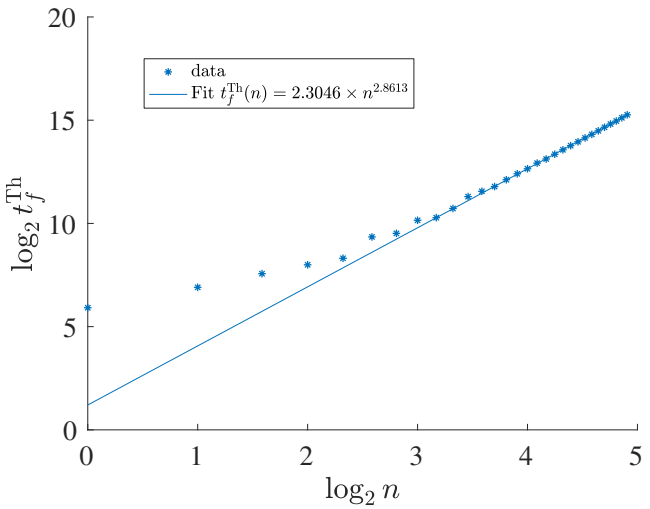


FIG. 3. The minimum time required to reach a success probability of $p_{\mathrm{Th}} = 0.95$ as a function of size size n for the noiseless quantum annealing glued trees algorithm. The solid line corresponds to a scaling of $n^{2.8613}$.

Since the scaling $\mathcal{O}(n^6)$ is an analytically derived upper bound, we expect and find the scaling obtained via numerical simulations to be better. To see this, let us define the threshold annealing time to be the minimum time required for the success probability (where success is defined as reaching the EXIT vertex) to reach a threshold probability p_{Th}:

$$t_f^{\mathrm{Th}}(n) \equiv \min\{t_f : p_{\mathrm{GS}}(t_f) \geq p_{\mathrm{Th}}\}, \tag{9}$$

(henceforth, we choose $p_{\mathrm{Th}} = 0.95$). In Fig. 3, we plot the scaling of $t_f^{\mathrm{Th}}(n)$ for the QA algorithm for the glued trees problem. The scaling is $\mathcal{O}(n^{2.86})$, which is significantly faster than $\mathcal{O}(n^6)$.

It is instructive to examine the $p_{\mathrm{GS}}(t_f)$ function. This is exhibited for the case $n = 10$ in Fig. 4. For $n = 10$, the threshold timescale is $t_f^{\mathrm{Th}}(10) = 1690$. This corresponds to the second peak in the oscillations. In general, the QA algorithm works by being in a region of the $p_{\mathrm{GS}}(t_f)$ function before adiabaticity is achieved.

It is also instructive to examine what the dynamics look like at different evolution timescales. We examine the populations in the instantaneous ground state, first excited state, and the second excited state as a function of the interpolation parameter s for $n = 4, 20$ in Fig. 5. For $n = 4$, at relatively small annealing times the evolution is close to optimal: the population starts off in the ground state, enters the first excited state at the first exponentially small gap, and returns to the ground state at the second exponentially small gap. At longer annealing times the dynamics is closer to adiabatic, with some interesting fluctuations that arise around the exponentially small gaps. For $n = 20$, at the threshold annealing time, the evolution is optimal and exhibits sharp transitions.

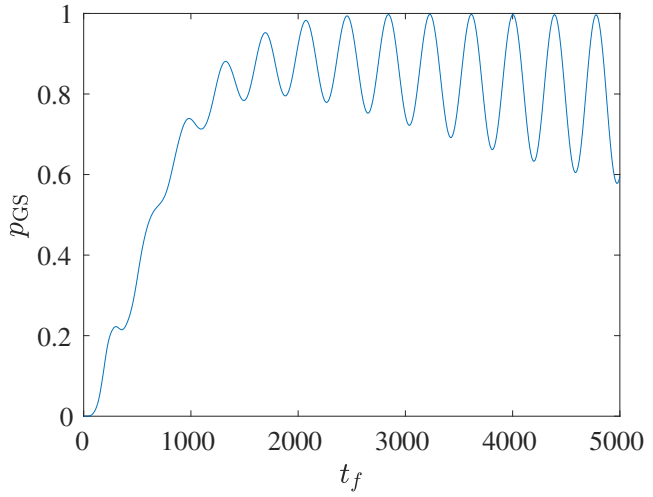


FIG. 4. Probability of finding the ground state at the end of evolution as a function of t_f for the noiseless glued-trees quantum anneal at problem size $n = 10$. Strong oscillations are observed, indicating that the optimal time at which we should terminate the algorithm is sub-adiabatic.

III. NOISE

The noise models we consider here are phenomenological. They ignore the details of how the noise may be realized and instead posit some general properties that noisy systems might have. This method is especially well-suited to oracle algorithms. This is because oracles are typically very difficult to realize physically. Indeed, for the glued-trees problem we can show that the terms H_0, H_1, and A in Eq. (1) all need to be highly nonlocal and require experimentally difficult-to-engineer interac-

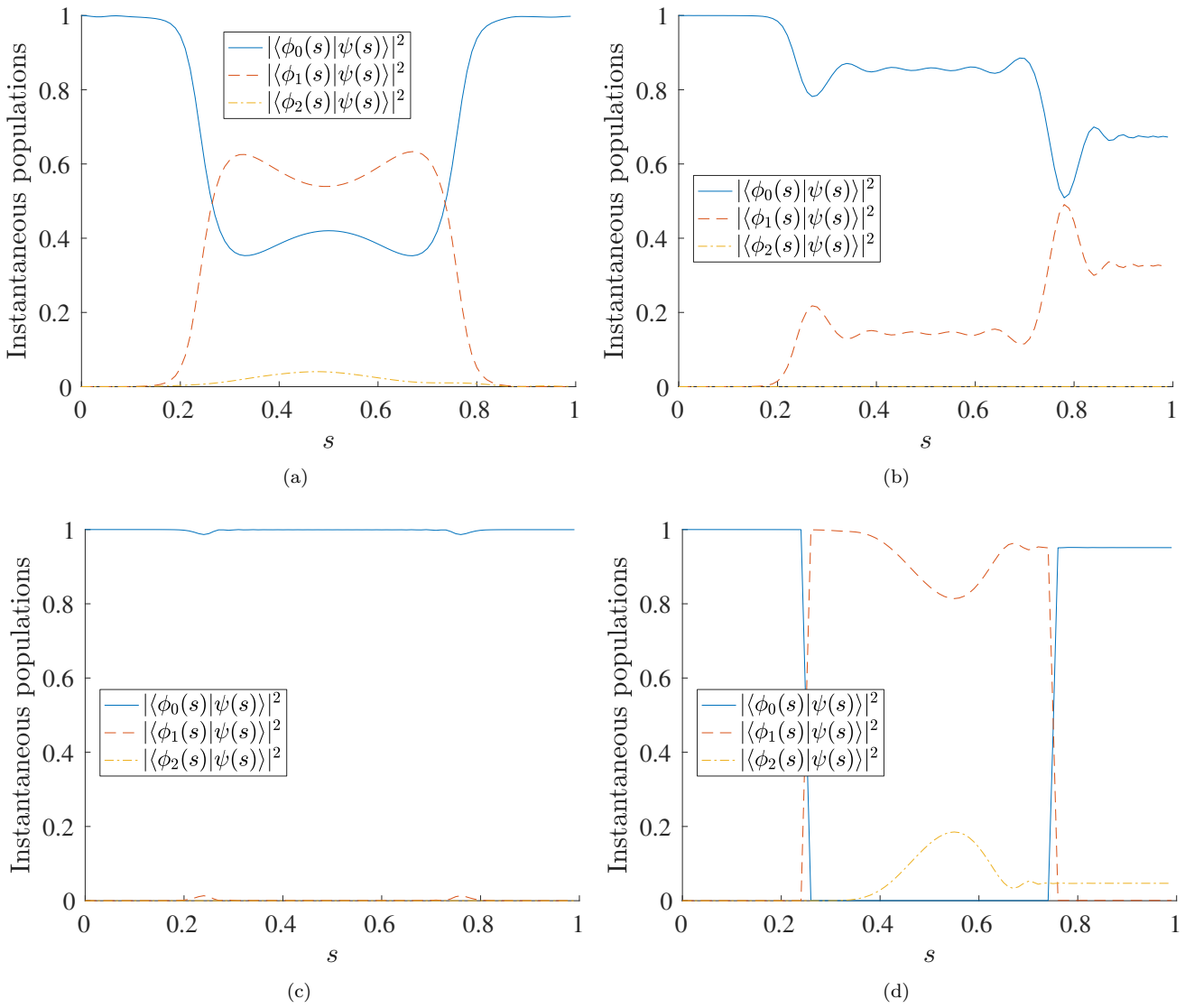


FIG. 5. (Color online) Populations in the instantaneous ground state, first excited state, and second excited state as a function of the anneal parameter s for the glued-trees problem without any added noise. (a) $n = 4$, $t_f = 268$; (b): $n = 4$, $t_f = 1000$; (c): $n = 4$, $t_f = 4000$; (d) $n = 20$ at $t_f = t_f^{\mathrm{Th}}(20) = 12125$ [see Eq. (9)].

tions (see Appendix B). Such oracle-level noise models are studied, e.g., in Refs. [17–19] for circuit algorithms, for the adiabatic Grover algorithm [20], and for some quantum walk algorithms (see [21] for a review), including the quantum walk version of the glued-trees algorithm [22].

Our noise model is inspired by one due to Roland and Cerf [23]. The noise model they consider is a time-dependent random-matrix added to the Hamiltonian, with the entries of the random matrix being time-dependent random variables distributed as white noise with a cut-off. They show that this noise does not significantly affect the performance of the adiabatic algorithm as long as the cut-off frequency of the white noise is either much smaller or much greater than the energy scale of the noiseless Hamiltonian. They further explore this noise model in detail for the adiabatic Grover algorithm [24].

The noise models we study are as follows. We add a time-independent random matrix h to our Hamiltonian $H(s)$ [Eq. (1)]. We write our noisy Hamiltonian $\tilde{H}(s)$ as

$$\tilde{H}(s) = H(s) + \epsilon h \ . \tag{10}$$

We restrict the noise matrix h to be inside the subspace spanned by the column basis, i.e., h has the same dimensions as $H(s)$ when written in the column basis. The four noise models we consider correspond to different ways of choosing the random matrix h.

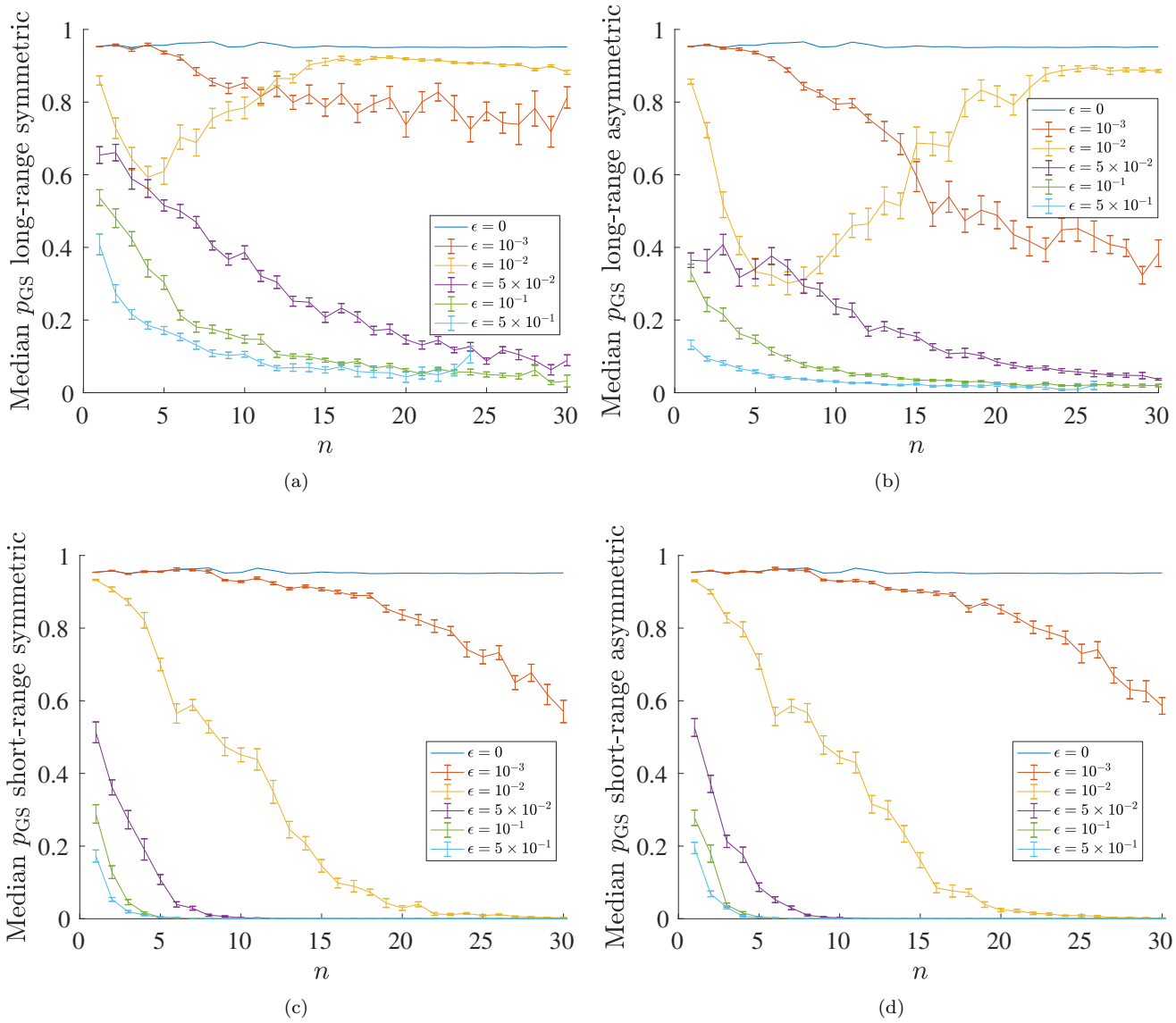


FIG. 6. (Color online) The median success probability, $p_{\rm GS}$, at the end of an evolution of duration $t_f^{\rm Th}(n)$ as a function of n for $\epsilon = 0, 10^{-3}, 10^{-2}, 5\times10^{-2}, 10^{-1}, 5\times10^{-1}$. $\epsilon = 0$ is the noiseless evolution (ϵ increases from top to bottom at $n = 1$). $t_f^{\rm Th}(n)$ is chosen so that the success probability for the noiseless probability is just above 0.95. The error bars are obtained by bootstrap sampling over 300 realizations of the noise. (a) The long-range symmetric noise model. (b) The long-range asymmetric noise model. (c) The short-range symmetric noise model. (d) The short-range asymmetric noise model.

A. Four noise models

We construct four noise models by selecting one branch in each of two dichotomies. The first dichotomy is between long-range and short-range noise models. The second dichotomy is between reflection-symmetric and reflection-asymmetric noise models.

First we consider a noise model in which h is chosen from the Gaussian Orthogonal Ensemble (GOE). This means that in any given basis, and in particular the column basis, the matrix elements of h are distributed as

$$h_{ij} = h_{ji} = \begin{cases} \mathcal{N}(0,1), & i \neq j \\ \mathcal{N}(0,2), & i = j \end{cases} . \tag{11}$$

(Here Gaussian random variables are denoted as $\mathcal{N}(\mu, \sigma^2)$, with μ being the mean of the Gaussian and σ the standard deviation.) A standard way of generating such a matrix is to start with a matrix M whose entries are independent $\mathcal{N}(0,1)$ random variables (and therefore

non-Hermitian) and setting

$$h = \frac{M + M^T}{\sqrt{2}}. \tag{12}$$

That Eq. (11) is obtained from Eq. (12) can be seen from the fact that the addition of two independent Gaussian random variables obeys

$$\mathcal{N}(\mu_1, \sigma_1^2) \pm \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2), \tag{13}$$

combined with $a\mathcal{N}(\mu, \sigma^2) = \mathcal{N}\left(\mu, (a\sigma)^2\right)$ (see Appendix C for more details about the GOE).

We call this noise model—i.e., the model generated by adding a time-independent random matrix chosen from the GOE—the *long-range asymmetric* (LA) noise model and denote h in this case by h_{LA}. *Long-range* refers to the fact that h contains matrix elements which connect all columns to all columns, and *asymmetric* refers to h breaking the reflection symmetry of the Hamiltonian. To see that h breaks the reflection symmetry, notice that h is not invariant under conjugation with the permutation matrix P, which, together with the fact that h is time-independent, yields that $\tilde{H}$ is not reflection symmetric.

Next, we consider what we call the *long-range symmetric* (LS) noise model. This noise model preserves the reflection symmetry of the problem. To generate this noise model, we first pick a matrix ω from the GOE. Then we reflection-symmetrize it:

$$h_{\mathrm{LS}} \equiv \frac{\omega + P\omega P}{\sqrt{2}}. \tag{14}$$

Now h_{LS} is manifestly reflection symmetric and therefore so is $\tilde{H} = H + \epsilon h_{\mathrm{LS}}$. Note that h_{LS} still contains terms connecting distant columns. From the above definition, we can check that the matrix elements of h_{LS} are distributed as

$$\begin{aligned}(h_{\mathrm{LS}})_{ij} &= (h_{\mathrm{LS}})_{ji} \\ &= (h_{\mathrm{LS}})_{(2n+1)-i,(2n+1)-j} = (h_{\mathrm{LS}})_{(2n+1)-j,(2n+1)-i} \\ &= \begin{cases} \mathcal{N}(0,1), & i \neq j \\ \mathcal{N}(0,2), & i = j. \end{cases}\end{aligned} \tag{15}$$

The reflection symmetry of h_{LS} implies that the spectrum of $\tilde{H}$ is reflection symmetric as well in this case.

We next turn to the short-range noise models. These noise-models only connect neighboring columns in the glued-trees graph. We examine both *short-range asymmetric* (SA) and *short-range symmetric* (SS) noise models. In the SA model, the Hamiltonian perturbation is given by

$$(h_{\mathrm{SA}})_{ij} = \begin{cases} \mathcal{N}(0,1) & |i-j| = 1, \\ 0 \quad \text{otherwise.} \end{cases} \tag{16}$$

In the SS model, we have

$$h_{\mathrm{SS}} = \frac{h_{\mathrm{SA}} + P h_{\mathrm{SA}} P}{\sqrt{2}}, \tag{17}$$

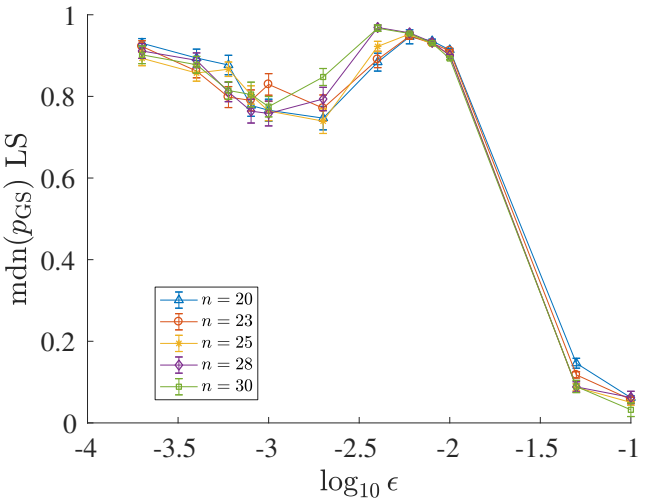


FIG. 7. (Color online) Median success probability vs. the log (base 10) of the strength of the noise for the LS model for several, larger values of the problem size n. There is a fall, then a rise, and then again a fall in this success probability as a function of ϵ for all values of n displayed.

which preserves the reflection symmetry of the Hamiltonian.

We remark that the parameter controlling the strength of the noise ϵ, can be absorbed into the standard deviations of the Gaussian random variables: e.g., $\epsilon\mathcal{N}(0,1) = \mathcal{N}(0,\epsilon^2)$. Therefore, the larger the noise, the greater the spread of the Gaussians according to which the matrix elements are drawn.

IV. NOISY GLUED-TREES: RESULTS FROM NUMERICAL SIMULATIONS

We simulate the Schrödinger evolution

$$i\hbar\frac{d}{dt}|\tilde{\psi}_{t_f}(s)\rangle = t_f\tilde{H}(s)\,|\tilde{\psi}_{t_f}(s)\rangle \tag{18}$$

using the different noise models and calculate the probability of finding the EXIT vertex at the end of the anneal, i.e., the success probability:

$$p_{\mathrm{GS}}[t_f^{\mathrm{Th}}(n)] \equiv \left|\langle\phi_0(s)|\tilde{\psi}_{t_f^{\mathrm{Th}}(n)}(1)\rangle\right|^2. \tag{19}$$

We choose the annealing timescale t_f to be equal to threshold timescale of the noiseless algorithm, i.e., the timescales depicted in Fig. 3. We do this because it is natural to imagine that one operates the algorithm in the regime in which the noiseless algorithm succeeds. Note that this probability is a random variable whose value depends on the specific noise realization, and we focus on the typical (median) value of this random variable. The median p_{GS} will depend on (a) the noise model, (b) the strength of the noise ϵ, and (c) the problem size n.

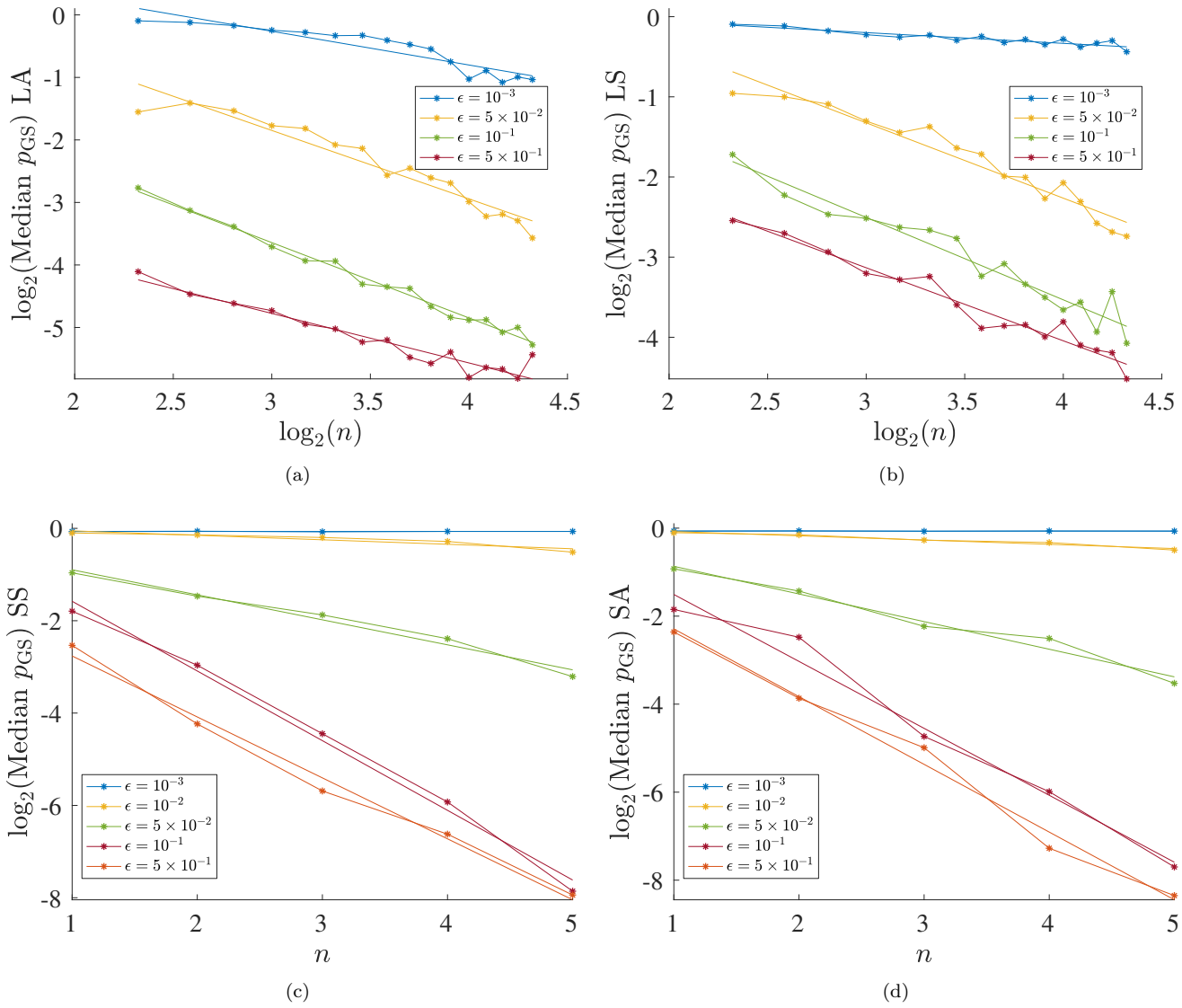


FIG. 8. (Color online) Base-2 logarithm of the median ground state success probability, as a function of problem size n, or the logarithm of the problem size $\log_2 n$, at different noise levels ϵ. Polynomial fits of the form $\mathcal{O}(n^\alpha)$ are performed for the long-range models. Exponential fits of the form $\mathcal{O}(2^{\alpha n})$ are performed for the short range models. This is done for $\epsilon \in \{10^{-3}, 10^{-2}, 5\times10^{-2}, 10^{-1}, 5\times10^{-1}\}$ for the short-range models (ϵ increases from top to bottom at $n = 1$). For the long-range models, the set of values of ϵ is the same as for the short-range models, except that we omit the $\epsilon = 10^{-2}$ case since it shows anomalous behavior (i.e., a rise and a fall) which doesn't fit a decay. The scaling coefficient α as a function of the noise ϵ is shown in Fig. 9. (a) The long-range symmetric noise model. (b) The long-range asymmetric noise model. (c) The short-range symmetric noise model. (d) The short-range asymmetric noise model.

In Fig. 6, we plot the median success probability for the four different noise models specified in Sec. III A as a function of problem size n, equal to the depth of one of the two binary trees.

The first observation is that the median p_{GS} behavior is not significantly different between the symmetric and asymmetric noise model, for both the long-range and short-range variants, although the symmetric noise does slightly outperform the asymmetric case. This suggests that while the symmetry of the spectrum is an important aspect of the performance of the noiseless algorithm (by allowing transitions from and back to the ground state), the noisy algorithm is somewhat robust to reflection symmetry-breaking.

Next, a remarkable feature of Fig. 6 is that for large enough n ($n \gtrsim 13$), the success probability is non-monotonic with respect to the strength of the noise ϵ. This can be seen more clearly in Fig. 7: as expected, there is a fall in the success probability from $\epsilon = 0$ to $\epsilon = 10^{-3}$, but, surprisingly, there is a rise from $\epsilon = 10^{-3}$ to $\epsilon = 10^{-2}$. For higher values of ϵ, the probability falls off, again as expected. We explain the counterintuitive

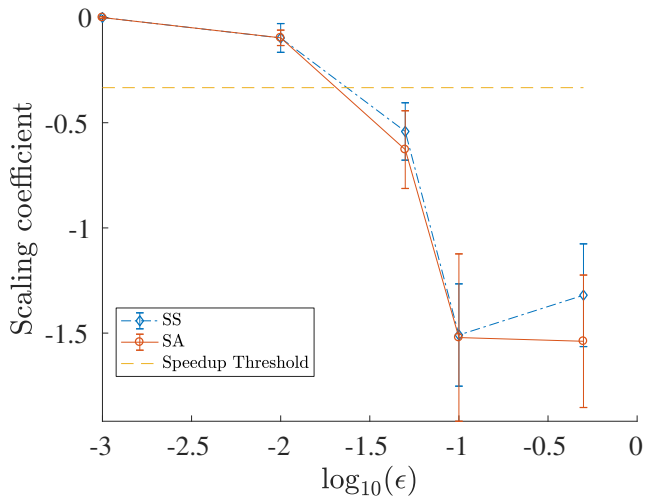


FIG. 9. (Color online) The exponential scaling coefficient α as a function of the base-10 logarithm of the strength of the noise ϵ for the short-range noise models. (The long-range noise models do not show an exponential scaling.) The scaling coefficient α is obtained by performing an exponential fit of the form $\mathcal{O}(2^{\alpha n})$, to the median success probability, $p_{\mathrm{GS}}(t_f^{\mathrm{Th}})$ vs. n curves [shown in Figs. 8(c) and 8(d)]. The dashed horizontal line at $\alpha = -1/3$ represents the scaling coefficient below which the speedup, over the best possible classical algorithm to solve the glued-trees problem, is lost. Thus, the short-range models lose the speedup for $\epsilon \gtrsim 10^{-1.75}$. The error-bars are 95% confidence intervals. While some error-bars are large due to the fits being performed on a small number of data points, the trend in the data is clear.

rise in the next section.

We also examine, for a given noise model and a given noise strength, whether the quantum speedup of the noiseless algorithm is retained. Since $2^{n/3}$ is the best possible time scaling that a noiseless classical random walk can achieve (see Sec. II and Ref. [9]), the $p_{\mathrm{GS}}(t_f^{\mathrm{Th}})$ vs. n scaling should decline no faster than $2^{-n/3}$ for a quantum speedup to persist. But, it might not be an exponential speedup: if the success probability for the noisy quantum algorithm declines as an exponential function that decreases slower than $2^{-n/3}$, then the speedup over the classical algorithm will instead be a polynomial speedup.

We thus perform fits to the $p_{\mathrm{GS}}(t_f^{\mathrm{Th}})$ vs. n curves, displayed in Fig. 8. For the long-range models, we exclude small values of n because the asymptotic behavior starts to show only at larger values of n. For the short-range models, we only perform the fits on small values of n because exponential decay causes the values of the success probability to be very small for larger values of n. The long-range models fit polynomials of the form $\mathcal{O}(n^{\alpha})$. This means that the long-range models have an exponential speedup over the noiseless classical algorithm. (We argue below why in fact referring to this as a quantum speedup is misleading.) On the other hand, the short-range models fit exponentials of the form $\mathcal{O}(2^{\alpha n})$, which means that they do not exhibit a polynomial speedup over the noiseless classical algorithm if $\alpha < -1/3$ (as discussed in the previous paragraph). In Fig. 9, we plot the scaling coefficient α as a function of the noise strength ϵ for the short-range noise models. As is apparent from Fig. 9, the quantum speedup is lost for the short-range models for $\epsilon \gtrsim 10^{-1.75} \approx 0.018$.

Let us now explain why the exponential speedup exhibited by the long-range noise models is misleading. For this speedup to count as a genuine quantum speedup, we must compare the quantum algorithm with an appropriate classical algorithm, so that we do not bias our analysis in favor of the quantum algorithm. How do we construct the appropriate classical algorithm in this case? Recall that in the quantum case, the long-range noise Hamiltonians have $\mathcal{N}(0,1)$-distributed off-diagonal terms. These terms connect distant columns of the graph. Thus, these models ought to be compared with classical random walks which have $\mathcal{N}(0,1)$-distributed transition probabilities between distant nodes of the graph. This needs to be normalized by a factor that is $\mathcal{O}(n)$ since there are $2n+2$ columns in total. Therefore, at any given time-step, in whichever column the random walker is, there is an $\mathcal{O}(1/n)$ probability that the random walker will transition directly to the EXIT vertex in the next time-step. Hence such a classical random walk will land at the EXIT vertex in $\mathcal{O}(n)$ time, and compared to the appropriate classical algorithm, the quantum algorithm with the long-range noise does not have an exponential speedup.

V. NOISE-INDUCED ADIABATICITY

Two things stand out in the results presented in the previous section. The first is that for the long-range models, there is a rise in success probability from $\epsilon \approx 10^{-3}$ to $\epsilon \approx 10^{-2}$. The second is that the long-range models have an exponential speedup over the noiseless classical algorithm for a wide range of noise strengths, when compared to the short-range models. In this section, we will provide explanations for these two observations.

One phenomenon helps explain both these behaviors: In the long-range models, the noise typically leads to a larger spectral gap. To see this, consider what happens to the spectral gap at first-order in perturbation theory when we add long-range noise. Consider first the asymmetric case. Let $E_0^{(1)}(s)$ and $E_1^{(1)}(s)$ be the first-order corrections to the ground- and first excited state respectively. We know that

$$E_0^{(1)}(s) = \epsilon \langle \phi_0(s) | h_{\mathrm{LA}}(s) | \phi_0(s) \rangle \tag{20a}$$
$$\sim \epsilon \mathcal{N}(0,2), \tag{20b}$$

where $|\phi_0(s)\rangle$ represents the ground state of the unperturbed (noiseless) problem at s. We have used the fact that $h_{\mathrm{LA}}(s)$ is drawn from the GOE, and that the diagonal elements of a GOE matrix are distributed with variance 2 [see Eq. (11)]. A similar calculation gives that

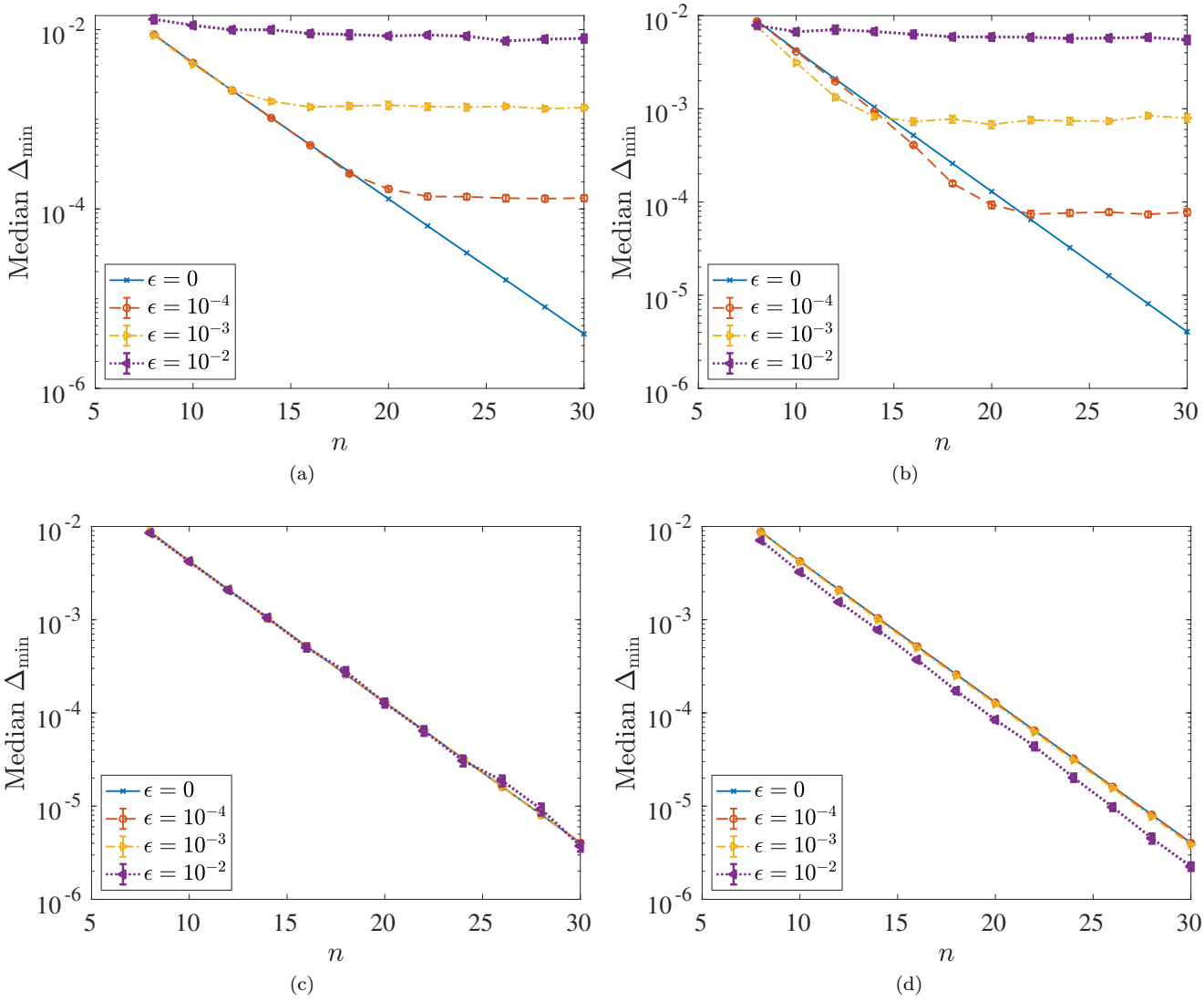


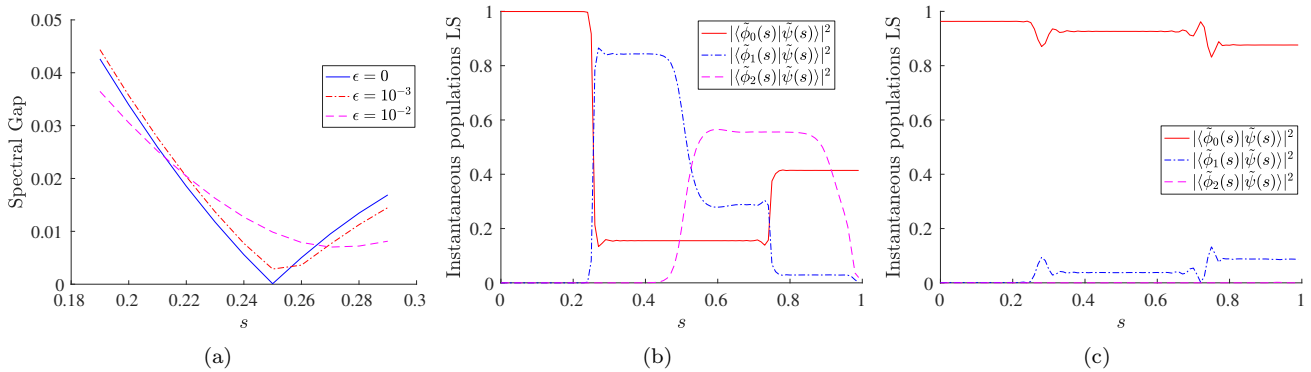
FIG. 10. (Color online) The scaling of the median minimum gap (log-scale) as a function of problem size n for the four noise models with $\epsilon \in \{10^{-4}, 10^{-3}, 10^{-2}\}$. The solid blue line represents the noiseless algorithm ($\epsilon = 0$), which has an exponentially closing minimum gap. (a) Long-range symmetric. (b) Long-range asymmetric. (c) Short-range symmetric. (d) Short-range asymmetric. The long-range models display a constant gap with problem size, while the short-range models exhibit an exponentially decreasing gap with problem size. This is consistent with the perturbative argument given in Eq. (21).

$E_1^{(1)}(s) \sim \epsilon\mathcal{N}(0,2)$. Putting these together, we can approximate the spectral gap of the perturbed Hamiltonian using first-order perturbation theory as follows.

$$\tilde{\Delta}_{\mathrm{LA}}(s) = \tilde{E}_1(s) - \tilde{E}_0(s) \tag{21a}$$
$$\approx [E_1(s) + \epsilon E_1^{(1)}(s)] - [E_0(s) + \epsilon E_0^{(1)}(s)] \tag{21b}$$
$$= \Delta(s) + \epsilon\mathcal{N}(0,4), \tag{21c}$$

where $\Delta(s)$ is the spectral gap of the noiseless problem. To obtain the last line we used Eq. (13) again. Using the fact that $\Delta(s)$ scales either as inverse polynomially or inverse exponentially (shown in Ref. [10]; see also Fig. 2), and the fact that the random variable $\mathcal{N}(0,4)$ has no scaling with problem size we can conclude that, typically, at first order in perturbation theory, the perturbed problem has an $\mathcal{O}(1)$ gap. (A similar argument establishes that the gap is $\mathcal{O}(1)$ for the case of the LS model as well.)

This argument will only work if $\epsilon\mathcal{N}(0,4)$ does not make the right-hand side of Eq. (21c) negative; if the right-hand side is negative, the perturbative approximation breaks down. The RHS is distributed according to $\mathcal{N}(\Delta(s), 4\epsilon^2)$. This is a Gaussian centered around a positive mean, and thus the chance of this distribution sampling negative values is small for small ϵ. So, heuristically, we expect the perturbative argument to work in typical instances. To corroborate the conclusion of this argument, we display the scaling of the gaps with problem size in Fig. 10. It is apparent that the long-range models



(a) (b) (c)

FIG. 11. (Color online) Two random instances of the long-range symmetric noise for $n = 20$ and $t_f = 12125$, with $\epsilon = 10^{-3}$ and $\epsilon = 10^{-2}$. (a) The spectral gap between the ground state and the first excited state for the two noisy instances and the noiseless case ($\epsilon = 0$), around $s = s^*$, i.e., near the location of the first exponentially closing gap in the noiseless problem. Both noisy instances increase the spectral gap from its value in the noiseless problem, with the increase being larger in the $\epsilon = 10^{-2}$ case. (b) The populations in the lowest three eigenstates of the noisy Hamiltonian as a function of the anneal parameter s, for $\epsilon = 10^{-3}$. The diabatic transitions do not happen as cleanly as in the noiseless case [shown in Fig. 5(d)]. (c) As in (b), for $\epsilon = 10^{-2}$. The dynamics are very close to adiabatic.

exhibit a constant scaling with problem size, while the short-range models exhibit an exponential scaling.

Note that the perturbative argument presented above is not directly applicable for the short-range noise models, because for these models matrix elements that are arbitrarily far apart in the column basis are not normally distributed.

Let us see how the perturbative lifting of the spectral gap helps explain the non-monotonic dependence on noise strength seen in Fig. 7. For $\epsilon \approx 0$, the algorithm succeeds because of the diabatic transitions from the ground state to the first excited state and then back down to the ground state; recall Figs. 5(a) and 5(d). As we increase ϵ from zero, the slight lifting of the gap interferes with these diabatic transitions, leading to a somewhat smaller success probability. This corresponds to the local minimum in Fig. 7. As we increase ϵ further, the gap increases more and this causes the dynamics to turn adiabatic, which increases the success probability. This corresponds to the second peak in Fig. 7. These two effects can be seen in Fig. 11, which shows typical instances of the noisy spectrum and the noisy dynamics under the LS noise model, at $\epsilon = 10^{-3}$ and $\epsilon = 10^{-2}$, for $n = 20$. Figure 11(a) shows that both noise realizations increase the spectral gap from its value in the noiseless case, more so for the higher ϵ realization. Figure 11(b) shows how the diabatic transitions are scrambled due to the noise. Then, as we increase the noise to $\epsilon = 10^{-2}$ in Fig. 11(c), we observe the onset of adiabaticity.

As we increase ϵ to values greater than 10^{-2}, the success probability falls off because even if the dynamics are adiabatic, the noisy spectrum and eigenstates have little relationship with the noiseless spectrum and eigenstates. To corroborate this, in Appendix D, we show using perturbation theory that the overlap between the unperturbed ground state and the perturbed ground state decays as $1 - \mathcal{O}(\epsilon^2)$, which suggests that as we increase ϵ, even if the dynamics are adiabatic, the ground state found at the end of the noisy evolution has low overlap with the EXIT vertex.

The perturbative lifting of the gap also explains why the long-range models exhibit an exponential quantum speedup. Indeed, because the noise induces adiabaticity for a certain range of values of ϵ, then as long as the overlap between the unperturbed ground state and the perturbed ground state remains significant, the noisy quantum system can still solve the problem by evolving adiabatically as long as the anneal timescale is greater than the adiabatic timescale. The latter is given by the inverse of the perturbed gap squared, which is $\mathcal{O}(1)$, multiplied the norm of the Hamiltonian, which is $\mathcal{O}(\text{poly}(n))$. The anneal timescale is chosen such that it provides a speedup in the noiseless case [Eq. (9) represents one such choice], which means the noiseless dynamics are adiabatic relative to the polynomially small gap between the first and second excited states (see Sec. II B), and therefore, the long-range noise dynamics will be adiabatic relative the constant gap between the ground and first excited state.

A natural question arises at this stage. Is the anomalous behavior of success probability with increasing system simply due to the perturbation increasing the energy scale of the system? That is, because the long-range noise matrices are selected from the GOE (and variants thereof), the norm of the perturbation increases with system size as $\sqrt{n}$; one might think that since the energies in the system increase, then so do its energy gaps, and then the larger gaps are responsible for the increase in success probability.

To test this explanation, we checked what happens after normalization of the perturbation matrix by its largest eigenvalue. This normalization ensures that the

perturbation never adds an amount of energy that scales with system size. If it were true that the perturbative lifting of the gap is due to pumping energy into the system, we would expect the perturbative lifting to disappear upon normalization, and consequently also expect the anomalous behavior of success probability with increasing system size to disappear upon normalization. However, as seen in Fig. 12, the non-monotonic dependence of the success probability on n for intermediate ϵ values continues to hold even after normalization. (This will in turn lead to non-monotonic dependence of the success probability on ϵ for larger values of n, analogous to the behavior seen in Fig. 7.) This is qualitatively similar to the behavior seen in Fig. 6(b) for the long-range asymmetric noise model, where the perturbation matrix h is not normalized.

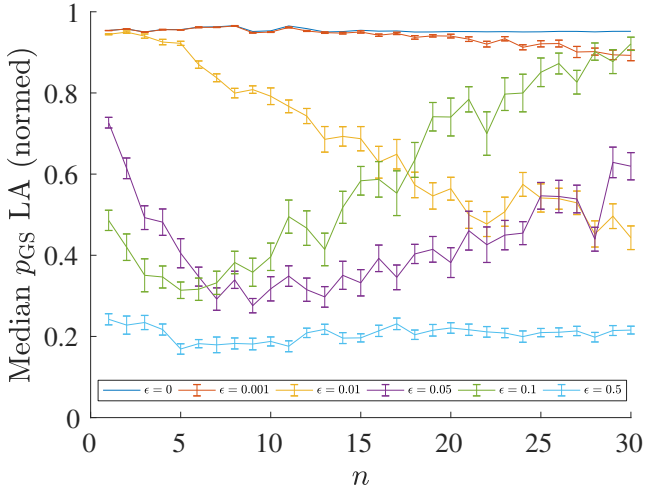


FIG. 12. (Color online) The median success probability, $p_{\rm GS}$, for the *normalized* long-range asymmetric noise case at the end of an evolution of duration $t_f^{\rm Th}(n)$, with ϵ increasing from top to bottom at $n = 1.t_f^{\rm Th}(n)$ is chosen so that the success probability for the noiseless probability is just above 0.95. Error bars were obtained by bootstrap sampling over 300 realizations of the noise.

VI. SUMMARY AND CONCLUSIONS

We have analyzed the quantum annealing algorithm for the glued trees problem under four different noise models: long-range vs. short-range and reflection-symmetric vs. reflection-asymmetric. These are oracular noise models: they add a Gaussian perturbation, of different forms, to the Hamiltonian evolution. We studied the success probability—i.e. the probability finding the EXIT vertex—at the end of the Schrödinger evolutions for these different noise models.

We found that the long-range noise models display a perturbative lifting of the spectral gap which causes the dynamics to transition from diabatic to adiabatic. This allows the algorithm subject to long-range noise models to solve the glued trees problem in polynomial time and hence display an exponential quantum speedup over the noiseless classical algorithm. This seems surprising, since it associates a robustness to noise with a quantum algorithm exhibiting exponential speedup. However, we argue that in fact this speedup is misleading, because it disappears when we compare the quantum algorithm to an appropriate classical analogue. More precisely, a classical random walk that has long-range transition probabilities will also be able to solve the problem in polynomial time and hence display an exponential speedup over the noiseless classical algorithm.

This analysis highlights that care must be taken in the selection of noise-models for oracular algorithms. Typically, oracles are hard to realize physically, so we must select phenomenological noise models for them. But in so choosing, we might end up with a model that changes the nature of the problem, which is what occurred in the long-range noise models. More precisely, the classical long-range noisy version of the algorithm changed its complexity from exponential into polynomial, so a quantum polynomially scaling algorithm for the problem cannot count as providing an exponential quantum speedup.

It is instructive to compare this with the results of Ref. [19], which analyzed the problem of learning the class of n-bit parity functions by making queries to a (noisy) quantum example oracle. There, the quantum algorithm has a linear speedup in the noiseless case, while it has a superpolynomial speedup when both the classical and quantum oracles are noisy. This happens upon depolarizing the qubits at the oracle's output at any constant nonzero rate.

For the glued trees problem we found a weaker result under the symmetric and asymmetric short-range noise models, which retain the exponential complexity of the classical problem. For sufficiently weak oracle noise, the quantum annealing algorithm retains a polynomial quantum speedup over the noiseless classical algorithm. But, for sufficiently strong oracle noise even the polynomial speedup is lost. The fact that for all values of the oracle noise the short-range noise models result in a loss of exponential speedup demonstrates that the exponential speedup of the glued-trees algorithm is not robust to noise.

We conjecture that more broadly, in the absence of fault tolerant error correction, exponential speedups cannot be obtained in any physical implementation of quantum annealing. This should not necessarily be a cause for pessimism: we are not ruling out polynomial speedups, which remain highly interesting and valuable.

ACKNOWLEDGMENTS

We are especially grateful to Hidetoshi Nishimori for an important observation regarding perturbative gap lifting. We also thank Milad Marvian and Evgeny Mozgunov

for useful discussions and Huo Chen and Richard Li for advice regarding parallel computation. The research is based upon work (partially) supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the U.S. Army Research Office contract W911NF-17-C-0050. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Computation for the work described in this paper was supported by the University of Southern California's Center for High-Performance Computing (hpc.usc.edu).

Appendix A: Hamiltonian in the column basis

Here we show how the matrix elements of the noiseless QA Hamiltonian are obtained in the column basis. I.e., we show how Eqs. (4) are obtained when Eq. (1) is written in the basis defined in Eq. (2).

H_0 and H_1 are straightforward: they are represented as $-|\mathrm{col}_0\rangle\langle\mathrm{col}_0|$ and $-|\mathrm{col}_{2n+1}\rangle\langle\mathrm{col}_{2n+1}|$ respectively. This immediately yields Eqs. (4a) and (4d). Consider the adjacency matrix A. Consider first $j < n$. In this case, in the column basis, $A_{j,j+1}$ is

$$\langle\mathrm{col}_j|A|\mathrm{col}_{j+1}\rangle = \langle\mathrm{col}_j|\left(\sum_{(x,x')\in E}|x\rangle\langle x'|\right)|\mathrm{col}_{j+1}\rangle \tag{A1}$$

$$= \frac{1}{\sqrt{N_j N_{j+1}}}\sum_{\substack{x\in\mathrm{col}_j; x'\in\mathrm{col}_{j+1}\\(x,x')\in E}} 1 \tag{A2}$$

$$= \frac{N_{j+1}}{\sqrt{N_j N_{j+1}}} \tag{A3}$$

$$= \sqrt{\frac{N_{j+1}}{N_j}} = \sqrt{\frac{2^{j+1}}{2^j}} = \sqrt{2}, \tag{A4}$$

where we have used the fact that for all columns to the left of the central glue, every vertex has exactly one edge connecting it to the column to its left. We have also used that for $j \le n$, $N_j = 2^j$. A parallel calculation will go through for $j > n+1$. Now, for $j = n$:

$$\langle\mathrm{col}_n|A|\mathrm{col}_{n+1}\rangle = \frac{1}{\sqrt{N_n N_{n+1}}}\times 2N_{n+1} \tag{A5}$$

$$= 2\times\sqrt{\frac{N_{n+1}}{N_n}} = 2, \tag{A6}$$

where we used that, at the central glue, there are two edges to every vertex. We have also used the fact that the number of vertices in the two columns at the glue is equal. Thus, we have:

$$\langle\mathrm{col}_j|A|\mathrm{col}_{j+1}\rangle = \begin{cases}2, & j=n\\ \sqrt{2}, & o.w.\end{cases}. \tag{A7}$$

For convenience, we redefine A such that $A_{\mathrm{new}} \equiv A_{\mathrm{old}}/\sqrt{2}$, giving us:

$$\langle\mathrm{col}_j|A|\mathrm{col}_{j+1}\rangle = \begin{cases}\sqrt{2}, & j=n\\ 1, & o.w.\end{cases}, \tag{A8}$$

which yields Eqs. (4b) and (4c).

Appendix B: Qubit versions of small glued trees instances

If we wanted to implement the glued-trees algorithm in a qubit system, we would need nonlocal and difficult-to-engineer interactions. We do not provide a mathematical proof of this claim, but the basic point can be illustrated using the case of $n = 1$, the smallest instance of the glued-trees problem which has 6 vertices. Consider the shortest possible naming system, in which each vertex is labelled by a length-3 bit-string. (Set aside, for the moment, the concern about this naming system interfering with the proof of classical hardness.) This means we can implement the QA Hamiltonian using 3 qubits.

We will name the vertices such that the Hamming distance between two vertices connected by an edge on the graph is as small as possible. This is done in an attempt to reduce the need for many-body terms as much as possible. For the case of $n = 1$, this is done as shown in Fig. 13.

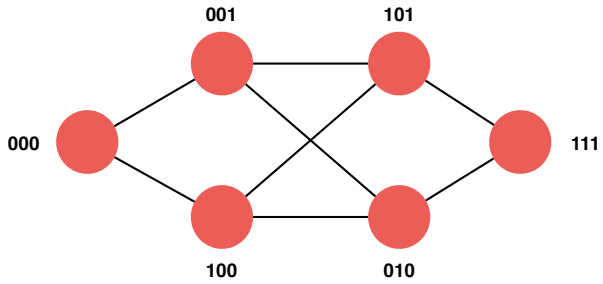


FIG. 13. (Color online) The smallest instance of the glued-trees problem. The graph is labelled with 3 bits. The naming is chosen so as to minimize the Hamming distance between vertices joined by an edge.

Clearly the adjacency matrix A of this graph is

$$A = |v_0\rangle\langle v_1| + |v_0\rangle\langle v_2| + |v_1\rangle\langle v_3| + |v_1\rangle\langle v_4| + |v_2\rangle\langle v_3| + |v_2\rangle\langle v_4| + |v_3\rangle\langle v_5| + |v_4\rangle\langle v_5| + h.c., \tag{B1}$$

where $v_0 \equiv 000$, $v_1 \equiv 001$, $v_2 \equiv 100$, $v_3 \equiv 101$, $v_4 \equiv 010$, and $v_5 \equiv 111$.

Let us convert this adjacency matrix into a local Hamiltonian. Let $\{I_i, X_i, Y_i, Z_i\}$ denote the Pauli operators on the i-th qubit ($i = 1, 2, 3$). We can compute $\mathrm{Tr}[AX_1]$, $\mathrm{Tr}[AX_1X_2]$, $\mathrm{Tr}[AX_1Y_2]$, and so on, for all the $4^3 = 64$ terms to obtain the coefficients of these terms in the Pauli representation of A. Doing this, we get

$$\begin{aligned} A =& \frac{1}{4}\,(2X_1 + X_2 + 2X_3 + X_1X_2 + X_2X_3 \\ &+ X_1X_3 + Y_1Y_2 + Y_2Y_3 \\ &- Y_1Y_3 + 2X_1Z_2 - X_2Z_3 - Z_1X_2 \\ &+ 2Z_2Y_3 - X_1Z_2X_3 + X_1X_2Z_3 \\ &+ Y_1Y_2Z_3 + Y_1Z_3Y_3 + Z_1X_2X_3 \\ &+ Z_1X_2Z_3 + Z_1Y_2Y_3)\,. \end{aligned} \tag{B2}$$

Note that there are no Z terms because A does not contain any diagonal matrix elements. We can already see that in order to implement this adjacency matrix we need 3-body interactions, and also "cross-term" interactions, i.e., interactions that couple, e.g., Z and X.

To complete the analysis, let us similarly write the Pauli representations of H_0 and H_1. This yields

$$\begin{aligned} H_0 =& \frac{1}{8}\,(Z_1 + Z_2 + Z_3 + Z_1Z_2 \\ &+ Z_2Z_3 + Z_1Z_3 + Z_1Z_2Z_3) \end{aligned} \tag{B3a}$$

$$\begin{aligned} H_1 =& \frac{1}{8}\,(-Z_1 - Z_2 - Z_3 + Z_1Z_2 + Z_2Z_3 \\ &+ Z_1Z_3 - Z_1Z_2Z_3)\,. \end{aligned} \tag{B3b}$$

Again 3-body interactions are required.

We have not ruled out the possibility that there exists an easily computed naming system for a general instance of the glued-trees problem, which generates a Hamiltonian representation for the QA Hamiltonian in such a way that the representation has small and constant locality, and the interactions required are simple (such as XX or YY). However, we are not aware of an explicit method for doing so and conjecture that none exists.

Appendix C: The Gaussian Orthogonal Ensemble

The GOE is the measure over the set of $N \times N$ real symmetric matrices described by

$$\Pr(h)dh = c_N \exp\left[-\frac{1}{2\sigma^2}\,\mathrm{Tr}(h^2)\right] \prod_i dh_{ii} \prod_{i<j} dh_{ij}, \tag{C1}$$

where σ is the so-called scale factor, c_N is a normalization, dh_{ii} and dh_{ij} are the standard Lebesgue measure. If $\{\lambda_i\}$ are the eigenvalues of h, then $\mathrm{Tr}(h^2) = \sum_i \lambda_i^2$. Therefore, the peak of this distribution is centered around matrices with eigenvalues close to 0. The GOE is invariant under conjugation by orthogonal matrices: i.e., if Eq. (C1) holds in one basis, then it also holds in another basis related by an orthogonal transformation to the first basis. See, e.g., Ref. [25] for more details.

Appendix D: Perturbative decay of overlap between the noisy and noiseless ground states

Here we calculate how the overlap between the noisy and noiseless groundstates decays as a function of ϵ for the case of the long-range noise models. This is done using perturbation theory, so we expect it to be valid for small ϵ.

Consider first the perturbative expansion for the perturbed ground state.

$$|\tilde{\phi}_0\rangle = c(\epsilon)\,|\phi_0\rangle + \epsilon \sum_{k>0} |\phi_k\rangle\,\frac{h_{k0}}{E_0 - E_k} + \mathcal{O}(\epsilon^2), \tag{D1}$$

Since $c(\epsilon) = \langle\phi_0|\tilde{\phi}_0\rangle$, this is the quantity we care about, namely, the overlap between the noisy and the noiseless ground states. Imposing $\langle\tilde{\phi}_0|\tilde{\phi}_0\rangle = 1$ (we are free to choose the normalization; see, e.g., Chapter 5 of Ref. [26]) and assuming $c(\epsilon) \in \mathbb{R}$, we get

$$c(\epsilon) \approx \left[1 - \epsilon^2 \sum_{k>0} \frac{h_{k0}^2}{(E_0 - E_k)^2}\right]^{\frac{1}{2}} \tag{D2a}$$

$$\approx 1 - \frac{\epsilon^2}{2} \sum_{k>0} \frac{h_{k0}^2}{(E_0 - E_k)^2}. \tag{D2b}$$

Thus the fidelity between the noisy and noiseless ground states goes as $1 - \mathcal{O}(\epsilon^2)$.

[1] T. Kadowaki and H. Nishimori, Phys. Rev. E **58**, 5355 (1998).

[2] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, Science **292**, 472 (2001).

[3] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Chemical Physics Letters **219**, 343 (1994).

[4] J. Brooke, D. Bitko, T. F., Rosenbaum, and G. Aeppli, Science **284**, 779 (1999).

[5] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, Science **295**, 2427 (2002).

[6] T. Albash and D. A. Lidar, Reviews of Modern Physics **90**, 015002 (2018).

[7] S. Jansen, M.-B. Ruskai, and R. Seiler, J. Math. Phys. **48**, 102111 (2007).

[8] A. Das and B. K. Chakrabarti, Rev. Mod. Phys. **80**, 1061 (2008).

[9] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing* (ACM, 2003) pp. 59–68.

[10] R. D. Somma, D. Nagaj, and M. Kieferová, Phys. Rev. Lett. **109**, 050501 (2012).

[11] D.R. Simon, in *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, edited by S. Goldwasser (IEEE Computer Society, Los Alamitos, CA, 1994) p. 116.

[12] P. W. Shor, *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on*, 35th Annual Symposium on Foundations of Computer Science, 1994 Proceedings , 124 (20-22 Nov 1994).

[13] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, 2002).

[14] An algorithm A has an exponential (polynomial) speedup over another algorithm B if the asymptotic scaling of A is an exponential (polynomial) function of the asymptotic scaling of B.

[15] That all the vertices, except the ENTRANCE and the EXIT vertices, have equal degree is crucial to avoid the easy solution of this problem by a backtracking classical random walk. See Ref. [9].

[16] The labeling scheme used for the vertices does not affect the performance of the quantum algorithm.

[17] N. Shenvi, K.R. Brown, and K.B. Whaley, Phys. Rev. A **68**, 052313 (2003).

[18] K. Temme, Physical Review A **90**, 022310 (2014).

[19] A. W. Cross, G. Smith, and J. A. Smolin, Physical Review A **92**, 012327 (2015).

[20] S. Mandrà, G. G. Guerreschi, and A. Aspuru-Guzik, Phys. Rev. A **92**, 062320 (2015).

[21] V. Kendon, Mathematical Structures in Computer Science **17**, 1169 (2007).

[22] J. Lockhart, C. Di Franco, and M. Paternostro, Physics Letters A **378**, 338 (2014).

[23] J. Roland and N. J. Cerf, Phys. Rev. A **71**, 032330 (2005).

[24] J. Roland and N. J. Cerf, Phys. Rev. A **65**, 042308 (2002).

[25] M. L. Mehta, *Random matrices*, 3rd ed., Vol. 142 (Elsevier, 2004).

[26] J. J. Sakurai and J. Napolitano, *Modern quantum mechanics* (Cambridge University Press, 2017).