

Error Propagation in NISQ Devices for Solving Classical Optimization Problems

Guillermo González-García^{1,2,*}, Rahul Trivedi^{1,2,†} and J. Ignacio Cirac^{1,2}

¹Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, Garching 85748, Germany

²Munich Center for Quantum Science and Technology (MCQST), Schellingstr. 4, Munich D-80799, Germany



(Received 13 April 2022; accepted 13 October 2022; published 5 December 2022)

We propose a random circuit model that attempts to capture the behavior of noisy intermediate-scale quantum devices when used for variationally solving classical optimization problems. Our model accounts for the propagation of arbitrary single-qubit errors through the circuit. We find that, even with a small noise rate, the quality of the obtained optima implies that a single-qubit error rate of $1/(nD)$ (where n is the number of qubits and D is the circuit depth) is needed for the possibility of a quantum advantage. We estimate that this translates to an error rate lower than 10^{-6} using the quantum approximate optimization algorithm for classical optimization problems with two-dimensional circuits.

DOI: [10.1103/PRXQuantum.3.040326](https://doi.org/10.1103/PRXQuantum.3.040326)

I. INTRODUCTION

Significant advances in the capabilities of quantum information processing hardware have recently been made, with the achievement of an important milestone of having reached quantum advantage [1–3]. In addition to developing technologies towards the final goal of a fault-tolerant quantum computer, there is widespread interest in exploring the capabilities of the currently available noisy intermediate-scale quantum (NISQ) devices [4]. This has inspired a number of heuristic quantum algorithms for NISQ devices [5–21], but it remains unclear if they can provide a quantum advantage for practically interesting problems.

One of the proposed applications for NISQ devices is to solve combinatorial optimization problems. Since this class of problems contains NP-hard instances, it is considered hard to solve on classical computers [22]. Quantum circuits can explore larger state spaces (e.g., entangled states) compared to their classical counterparts and hence there is a possibility of quantum speedup in some instances of these problems. Several heuristic variational quantum algorithms have been proposed for solving optimization problems [8–15], most notably the quantum approximate optimization algorithm (QAOA) [23]. While variational

algorithms in general lack provable guarantees for quantum speedups, there is a growing body of literature that suggests their use for practically interesting problems that remain hard to solve on classical computers [24–26], as well as provides evidence for their experimental feasibility [27,28]. However, a number of algorithmic challenges to these heuristics that stem from the limitations on the accessible circuit architectures, such as barren plateaus [29], expressibility of the ansatz [30,31], or reachability deficits [32], have been identified.

The influence of noise in quantum devices on variational algorithms is an important consideration for assessing their utility in the near term. Noise-induced barren plateaus in the optimization landscape of the variational ansatz have been identified [33]. Numerical modeling of the impact of noise has also been attempted [34,35], but this analysis is limited to very small circuit sizes. A set of rigorous results on the impact of noise in the quality of the optima obtained have recently been provided using entropic arguments [36–38]. By simply tracking the von Neumann entropy of the quantum state, in Refs. [37,38] it was argued that beyond circuit depths of $\Theta[\log(n)]$ with depolarizing noise, the output state is very close to the maximally mixed state. Stilck França and Garcia-Patron [36] improved on this result and showed that, even beyond circuit depths of $\Theta(1)$, the quality of the solution of a large class of classical optimization algorithms obtained from this circuit can be achieved with a classical algorithm.

However, while these analyses already provide provable limits on the quality of variational quantum algorithms that can inform current experiments, they likely underestimate the impact of noise in variational circuits used in practice. This underestimation arises from the fact that these bounds are applicable to *any* quantum circuit, and,

*guillermo.gonzalez@mpq.mpg.de

†rahul.trivedi@mpq.mpg.de

‡Both authors contributed equally.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

consequently, they also bound the performance of circuits that do not create significant entanglement in the quantum state. Therefore, these bounds do not capture propagation of errors through the quantum circuit. Furthermore, these bounds are also loose for noise models, such as amplitude damping noise, which can possibly *decrease* the entropy in the quantum circuit; however, for a typical quantum circuit, it is expected that the presence of such noise would still degrade its performance.

In this paper, we propose a random circuit model that attempts to analyze the impact of noise in NISQ devices used to variationally solve classical optimization problems. Unlike previously proposed random circuit models, our model captures the fact that these circuits map product states for product states, i.e., ideally, a typical member of this circuit family starts with a product state, builds entanglement in between the qubits, and then disentangles it to another (product) state that is the bit string that solves the classical optimization problem. Our study here is thus different from recent works that have studied the impact of noise in Haar-random quantum circuits [39–42] with geometric constraints, where a typical member of these circuit families has an anticoncentrated output state, which is in stark contrast to what is expected for variational quantum circuits that ideally output a state close to the ground state.

We perform an average-case analysis on this distribution of circuits, and find that noise propagates very rapidly through the circuit, severely limiting the performance of the circuits. We provide both numerical results and analytical scalings for three different architectures: one dimensional (1D), 2D, and nonlocal. Our results indicate that to obtain a solution to a classical optimization problem within a fixed multiplicative error of its true solution with a constant rate of noise p , the circuit depth of a typical member of this distribution has to be smaller than $\max(O(p^{-1/2}), O[1/(pn)])$ in the 1D architecture and $\max(O(p^{-1/3}), O[1/(pn)])$ in the 2D architecture. Furthermore, it follows from our analysis that the impact of error is qualitatively similar for different noise models in contrast to Ref. [36], which relies on the noise-channel being primitive and unital. Moreover, our analysis technique can be easily extended to understanding the impact of noise on variational quantum algorithms for solving the quantum optimization problem (i.e., finding the ground state of a quantum many-body Hamiltonian).

In addition to this average-case analysis, we provide a concentration result for low-depth local circuits that shows that this average is representative of a typical circuit. However, we note that this does leave open the question of whether a clever method of choosing the parameters of a variational quantum circuit could be devised (e.g., with a closed-loop optimization of the noisy quantum circuit), such that the resulting circuit avoids the predicted proliferation of noise.

II. ERROR PROPAGATION MODEL

Consider the problem of solving a classical optimization problem over $\{0, 1\}^n$, and suppose that the unitary circuit that maps an initial product state $|0\rangle^{\otimes n}$ to the solution $x^* \in \{0, 1\}^n$ is given by \mathcal{U}_{sol} . Since we are only interested in analyzing a lower bound on the impact of errors, we assume that the optimal circuit is already known and is noiseless, and do not address the problem of finding it. For depth D , we now consider a random quantum channel as $\Phi = \bigcirc_{t=1}^D T_t$, where the T_t are given by

$$T_t = \begin{cases} \mathcal{N} \circ \mathcal{U}^{(t)} & \text{if } t \leq D/2, \\ \mathcal{N} \circ [\mathcal{U}^{(D-t)}]^\dagger & \text{if } t > D/2, \end{cases} \quad (1)$$

where \mathcal{N} is a layer of the noise channel and $\mathcal{U}^{(t)}$ is a layer of two-qubit Haar random unitaries. The effective channel for the circuit under consideration, depicted in Fig. 1(a), will be constructed via $\Phi \circ \mathcal{U}_{\text{sol}}$.

Each member of the ensemble of circuits defined above solves the optimization problem in the absence of noise (i.e., when $\mathcal{N} = \text{id}$). Furthermore, in the absence of noise, the circuit builds entanglement in the input state (which is a product state) for the first $D/2$ layers (which we call the *entangling unitary*), after which it uncomputes these unitaries to finally obtain a product state. However, any error in the unitary circuit propagates and alters the final result. Since we are interested in the propagation of noise, we assume that the one-qubit gates are noiseless. We point out that a similar model can be constructed for quantum optimization problems, i.e., finding the ground state of a quantum many-body Hamiltonian, with \mathcal{U}_{sol} being the circuit that maps $|0\rangle^{\otimes n}$ to the ground state. Even while assuming \mathcal{U}_{sol} to be noiseless, the noisy entangling and unentangling unitaries will still capture propagation of errors within this model.

We consider three different architectures for generating $\mathcal{U}^{(t)}$ that model different interaction ranges that can be accessed on physical hardware [Fig. 1(b)].

- (a) A 1D local architecture with periodic boundary conditions, where alternating layers of nearest-neighbor gates are applied.
- (b) A 2D local architecture with periodic boundary conditions on a square lattice. Alternating layers of horizontal and vertical two-qubit gates are applied.
- (c) A nonlocal architecture where, for each layer, $n/2$ pairs are chosen at random, and $n/2$ two-qubit gates are applied between the pairs.

Given that, in the absence of noise, all instances of the random unitaries produce the same unitary, we consider the output of the channel averaged over the random unitaries

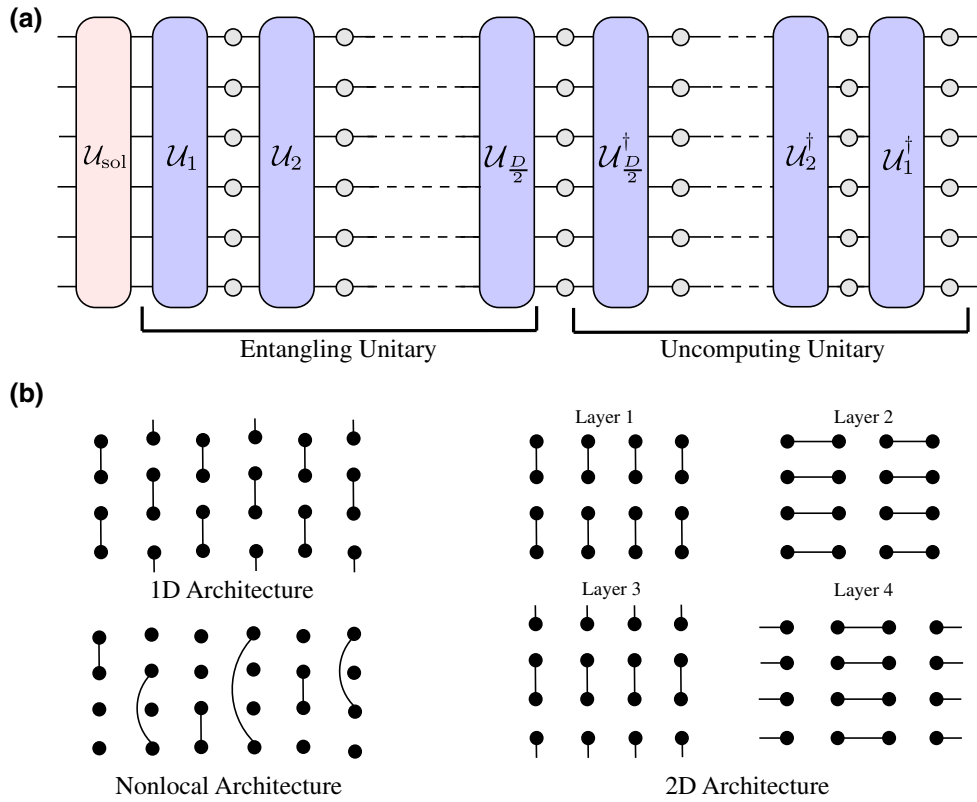


FIG. 1. (a) Schematic depiction of the ensemble of unitary circuits considered in this paper. The circuits consist of an entangling and uncomputing unitary that maps a product state to another product state in the absence of noise. (b) Different circuit architectures (1D, 2D, and nonlocal circuits) studied in this paper. All the circuit architectures are assumed to be translationally invariant—in particular, for the 1D and 2D architectures we assume periodic boundary conditions).

for each of the architectures,

$$\Phi_{\text{avg}}^{\mathcal{A}}(\rho) = \int_{\mathcal{U}} d\mathcal{U} \Phi(\rho) \quad \text{with } \mathcal{A} \in \{1\text{D}, 2\text{D}, \text{NL}\}, \quad (2)$$

where $\Phi_{\text{avg}}^{1\text{D}}$ will represent the averaged channel with the 1D architecture, $\Phi_{\text{avg}}^{2\text{D}}$ the 2D architecture, and $\Phi_{\text{avg}}^{\text{NL}}$ the nonlocal architecture [Fig. 1(b)].

An important fact that we use to calculate $\Phi_{\text{avg}}^{\mathcal{A}}$ is that the twirl of a two-qubit quantum channel \mathcal{M} over the Haar measure is a two-qubit depolarizing channel [Fig. 2(a)] [43], i.e.,

$$\begin{aligned} \mathcal{E}_{\text{dep}}(X) &= \int_{\mathcal{U}} d\mathcal{U} [\mathcal{U}^\dagger \mathcal{M} \mathcal{U}](X) \\ &= \lambda X + (1 - \lambda) \frac{I^{\otimes 2}}{4} \text{tr}(X), \end{aligned} \quad (3)$$

where if A_k represents the k th Kraus operator of the channel \mathcal{M} then

$$\lambda = \frac{1}{15} \left(\sum_k |\text{tr}(A_k)|^2 - 1 \right). \quad (4)$$

Since this expression applies to arbitrary noise channels, it follows that the specific choice of noise does not greatly impact the qualitative behavior of our model, with only the noise strength λ depending on the specific channel.

In the remainder of this paper, we consider the noise to be a depolarizing channel of strength p ,

$$\mathcal{N}(\rho) = (1 - p)\rho + p \text{tr}(\rho) I/2,$$

i.e., an error, interpreted as tracing out the qubit and replacing it with $I/2$, occurring with probability p . When this noise channel acts on two qubits, there are three distinct possibilities—no errors occurred (i.e., $\rho \rightarrow \rho$), both the qubits experienced errors ($\rho \rightarrow I^{\otimes 2}/4$), or only one qubit experienced error (i.e., $\rho \rightarrow I/2 \otimes \text{tr}_1(\rho)$ or $\text{tr}_2(\rho) \otimes I/2$). Then, by direct application of Eqs. (3) and (4), we can compute the effective two-qubit depolarizing obtained after averaging over the two-qubit unitaries for these three cases.

- (a) For $\mathcal{M}(\rho) = \rho$, i.e., no errors occur in either qubit between the application of \mathcal{U} and \mathcal{U}^\dagger ,

$$\mathcal{E}_{\text{dep}}(\rho) = \rho. \quad (5a)$$

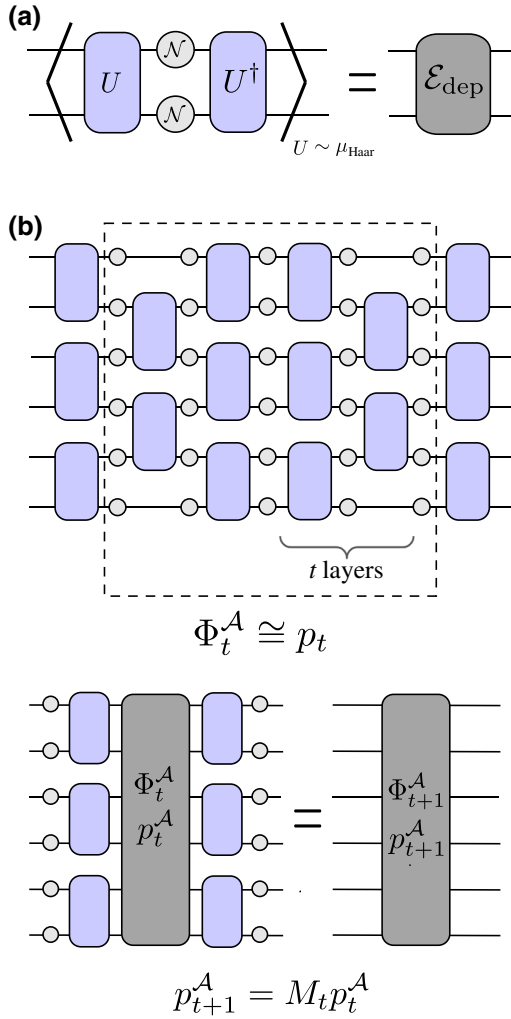


FIG. 2. (a) Averaging over Haar-random unitary U yields a two-qubit depolarizing channel \mathcal{E}_{dep} . (b) Schematic depiction of averaging over the circuit ensemble shown in Fig. 1—we average from the center of the circuit outwards, and derive an effective Markov chain that describes the channels obtained.

(b) If $\mathcal{M}(\rho) = \text{tr}(\rho)I^{\otimes 2}/4$, i.e., errors occurred in both qubits between the application of \mathcal{U} and \mathcal{U}^\dagger ,

$$\mathcal{E}_{\text{dep}}(\rho) = \text{tr}(\rho) \frac{I^{\otimes 2}}{4}. \quad (5b)$$

(c) If $\mathcal{M}(\rho) = I/2 \otimes \text{tr}_1(\rho)$ or $\text{tr}_2(\rho) \otimes I/2$, i.e., an error occurred in only one of the qubits in between the application of \mathcal{U} and \mathcal{U}^\dagger , then, after averaging over \mathcal{U} in Eq. (3),

$$\mathcal{E}_{\text{dep}}(\rho) = \frac{1}{5}\rho + \frac{4}{5}\text{tr}(\rho) \frac{I^{\otimes 2}}{4}. \quad (5c)$$

Physically, this can be interpreted as the error in one of the qubits propagating to the other qubit

with probability 4/5 due to the random two-qubit entangling gates.

Let us now consider the quantum channel Φ_{avg}^A ; the key idea to computing this average channel, depicted schematically in Fig. 2(b), is to start analyzing the circuit from its center (i.e., in between the entangling and uncomputing unitary layers), and iteratively construct the channel obtained on including additional unitary layers. We denote by Φ_t^A the channel formed by including and averaging over t layers of the entangling and uncomputing unitary from the center of the circuit. For $t = 0$, we have only a single layer of depolarizing noise and hence $\Phi_{t=0}^A = \mathcal{N}^{\otimes n}$; this channel applies the noise independently on each qubit or leaves their state unchanged. After including unitary layers and averaging over them, it follows from the analysis of the two-qubit errors shown above that the resulting channels still either apply noise on each qubit or leave them unchanged, although the noise is no longer applied independently on each qubit. Therefore, we assume the following ansatz for Φ_t^A :

$$\Phi_t^A = \sum_{\vec{j} \in \{0,1\}^n} p_t^A(\vec{j}) \left[\bigotimes_{\alpha=1}^n \tau_{j_\alpha} \right] \quad (6)$$

with

$$\tau_0(\cdot) = \text{id}(\cdot) \quad \text{and} \quad \tau_1(\cdot) = \text{tr}(\cdot) \frac{I}{2}.$$

Here, $p_t^A(\vec{j})$ for a given $\vec{j} \in \{0,1\}^n$ is the probability that the first qubit experiences noise if $j_0 = 1$ and remains unchanged otherwise, the second qubit experiences noise if $j_1 = 1$ and remains unchanged otherwise, and so on. We note that at $t = 0$, i.e., when $\Phi_{t=0}^A$ is a layer of depolarizing noise,

$$p_{t=0}^A(\vec{j}) = (1-p)^{n-|\vec{j}|} p^{|\vec{j}|}, \quad (7)$$

where $|\vec{j}| = \sum_{i=1}^n j_i$.

Next, we show that if Φ_t^A is of the form of Eq. (6) then so is Φ_{t+1}^A and we relate probability distribution p_{t+1}^A to p_t^A . Note that

$$\Phi_{t+1}^A = \mathbb{E}_{\mathcal{U}^{(t)}} [\mathcal{N}^{\otimes n} \mathcal{U}^{(t)\dagger} \Phi_t^A \mathcal{U}^{(t)} \mathcal{N}^{\otimes n}], \quad (8)$$

where $\mathcal{U}^{(t)}$ is the unitary applied at the t th layer from the center. Depending on the circuit architecture, $\mathcal{U}^{(t)} = \bigotimes_{(\alpha,\beta) \in S^{(t)}} \mathcal{U}_{\alpha,\beta}^{(t)}$, where $\mathcal{U}_{\alpha,\beta}^{(t)}$ is a two-qubit gate acting on qubits α, β and $S^{(t)}$ contains a list of qubits that interact with each other through the two-qubit unitaries. The averaging over each two-qubit unitary can now be done independently using Eq. (5). In particular, consider averaging over one of these unitaries acting on qubits α and β ;

from Eq. (6), it follows that

$$\begin{aligned} \mathbb{E}_{\mathcal{U}_{\alpha,\beta}^{(t)}} [\mathcal{U}_{\alpha,\beta}^{(t)\dagger} \Phi_t^A \mathcal{U}_{\alpha,\beta}^{(t)}] \\ = \sum_{\vec{j}} p_t^A(\vec{j}) \mathbb{E}_{\mathcal{U}_{\alpha,\beta}^{(t)}} \left[\mathcal{U}_{\alpha,\beta}^{(t)\dagger} \left(\bigotimes_{\gamma=1}^n \tau_{j_\gamma} \right) \mathcal{U}_{\alpha,\beta}^{(t)} \right] \\ = \sum_{\vec{j}} p_t^A(\vec{j}) \left(\bigotimes_{\gamma \neq \alpha, \beta} \tau_{j_\gamma} \right) \otimes \mathbb{E}_{\mathcal{U}_{\alpha,\beta}^{(t)}} [\mathcal{U}_{\alpha,\beta}^{(t)\dagger} \tau_{j_\alpha} \otimes \tau_{j_\beta} \mathcal{U}_{\alpha,\beta}^{(t)}]. \end{aligned}$$

where in the last step we have used the fact that channels $\mathcal{U}_{\alpha,\beta}^{(t)}, \mathcal{U}_{\alpha,\beta}^{(t)\dagger}$ commute with τ_{j_γ} for $\gamma \neq \{\alpha, \beta\}$ since they act only on qubits α, β . Now, from Eq. (5), it follows that

$$\begin{aligned} \mathbb{E}_{\mathcal{U}_{\alpha,\beta}^{(t)}} [\mathcal{U}_{\alpha,\beta}^{(t)\dagger} \tau_{j_\alpha} \otimes \tau_{j_\beta} \mathcal{U}_{\alpha,\beta}^{(t)}] \\ = \begin{cases} \tau_{j_\alpha} \otimes \tau_{j_\beta} & \text{if } j_\alpha = j_\beta, \\ \frac{1}{5} \tau_0 \otimes \tau_0 + \frac{4}{5} \tau_1 \otimes \tau_1 & \text{if } j_\alpha \neq j_\beta, \end{cases} \end{aligned}$$

and, therefore, it follows that

$$\mathbb{E}_{\mathcal{U}_{\alpha,\beta}^{(t)}} [\mathcal{U}_{\alpha,\beta}^{(t)\dagger} \Phi_t^A \mathcal{U}_{\alpha,\beta}^{(t)}] = \sum_{\vec{j}} q_{\alpha,\beta;t}(\vec{j}) \left[\bigotimes_{\alpha \in \vec{j}} \tau_\alpha \right],$$

where probability distribution $q_{\alpha,\beta;t}(\vec{j})$ is related to $p_t^A(\vec{j})$ via a transition matrix that is identity on all but the α th and β th bits, given by

$$\begin{aligned} M_{\alpha,\beta}^U[(j_\alpha, j_\beta) \rightarrow (j'_\alpha, j'_\beta)] \\ = \begin{cases} 1 & \text{if } (j_\alpha, j_\beta), (j'_\alpha, j'_\beta) = (0, 0), \\ 1/5 & \text{if } (j_\alpha, j_\beta) = (0, 1) \text{ or } (1, 0), (j'_\alpha, j'_\beta) = (0, 0), \\ 4/5 & \text{if } (j_\alpha, j_\beta) = (0, 1) \text{ or } (1, 0), (j'_\alpha, j'_\beta) = (1, 1), \\ 1 & \text{if } (j_\alpha, j_\beta), (j'_\alpha, j'_\beta) = (1, 1). \end{cases} \end{aligned} \quad (9)$$

Repeating this for all the two-qubit gates in unitary $\mathcal{U}^{(t)}$, we obtain

$$\mathbb{E}_{\mathcal{U}^{(t)}} [\mathcal{U}^{(t)\dagger} \Phi_t^A \mathcal{U}^{(t)}] = \sum_{\vec{j}} q_t^A(\vec{j}) \left[\bigotimes_{\alpha \in \vec{j}} \tau_\alpha \right],$$

where $q_t^A(\vec{j})$ is a probability distribution related to $p_t^A(\vec{j})$ via transition matrix M_t^U given by

$$M_t^U = \bigotimes_{(\alpha,\beta) \in S^{(t)}} M_{\alpha,\beta}^U.$$

Having averaged over the unitaries, we now consider applying the layers of depolarizing noise. This can be done

explicitly; from Eq. (8) we obtain

$$\Phi_{t+1}^A = \sum_{\vec{j} \in \{0,1\}^n} q_t^A(\vec{j}) \left[\bigotimes_{\alpha \in \vec{j}} (\mathcal{N} \tau_\alpha \mathcal{N}) \right].$$

Furthermore, it can be immediately seen from the definitions of the τ_α that

$$\mathcal{N} \tau_\alpha \mathcal{N} = \begin{cases} (1-p)^2 \tau_0 + (2p-p^2) \tau_1 & \text{if } \alpha = 0, \\ \tau_1 & \text{if } \alpha = 1. \end{cases}$$

Thus, probability distribution $p_{t+1}^A(\vec{j})$ is obtained from probability distribution $q_t^A(\vec{j})$ using transition matrix M^{noise} that, independently for every bit in \vec{j} , has transition probabilities

$$M^{\text{noise}}(j \rightarrow j') = \begin{cases} (1-p)^2 & \text{if } j = 0, j' = 0, \\ 2p-p^2 & \text{if } j = 0, j' = 1, \\ 1 & \text{if } j = 1, j' = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Combining the transition rules for both the unitary layer and the noise, we then obtain a Markov chain for probability distribution p_t^A ,

$$p_{t+1}^A = (M^{\text{noise}} M_t^U) p_t^A. \quad (11)$$

To analyze the state obtained at the output of this circuit, after averaging over all the unitaries, we thus only need to compute probabilities $p_{t=D/2}^A$ starting from initial distribution $p_{t=0}^A$ [Eq. (7)], which then gives us access to channel $\Phi_{t=D/2}^A$ (note that the effective time for which we need to evolve the Markov chain is half of the depth of the circuit, since each layer of the Markov chain accounts for a layer in the entangling and uncomputing unitaries). This can be computed by simulating the Markov chain with the transition matrices described above and allows us to analyze how many and which qubits in the solution of the optimization problem are, on average, correct. In the next section, we analyze this Markov chain both analytically and numerically to understand the impact of the noise and circuit depth on the quality of the output.

III. IMPACT OF ERRORS

A. Analysis of error propagation

We first compute the expectation value of the number of qubits that are depolarized at the end of the computation, which we denote as $\langle q \rangle$. This can be done numerically by sampling from the Markov chain described in the previous section using Markov chain Monte Carlo. The results are shown in Fig. 3—our results show that, as expected, all the

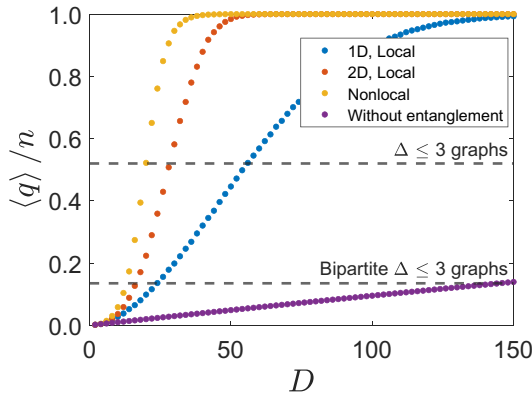


FIG. 3. We represent the expectation value of the fraction of depolarized qubits, $\langle q \rangle / n$, as a function of the circuit depth for the different architectures: 1D, 2D, and nonlocal. We use a system size $n = 900$ and an error rate $p = 10^{-3}$. Additionally, we represent $\langle q \rangle / n = 1 - (1 - p)^D$, which is the number of depolarized qubits that one would get when applying only local depolarizing noise to all the qubits, without any unitaries (and therefore without entanglement). We see that the convergence to uniform is very fast with our model. The horizontal lines represent thresholds for classical superiority for unweighted max-cut problems on arbitrary bounded degree-3 graphs and bipartite bounded degree-3 graphs: for values of $\langle q \rangle / n$ greater than the threshold, there is a classical efficient algorithm that outputs a better solution than the averaged quantum channel (see Sec. III B). The number of samples taken is 2000, which reduces the error in the estimated mean to 2%.

qubits depolarize exponentially fast with the circuit depths due to a rapid propagation of errors.

In order to better understand these numerical results, we would like to analyze theoretically the behavior of the Markov chain. We first provide heuristic estimates for the scalings of $\langle q \rangle$ with n . Note that every time a qubit is depolarized in the effective Markov chain, the neighboring qubit can be depolarized with a constant [i.e., $\Theta(1)$] probability. Take, for instance, the 1D model—at the end of the circuit, an error in the middle of the circuit will have propagated to $\min(O(D), n)$ qubits. We can now estimate the probability that a specific qubit will be depolarized at the end of the computation—on average, it will be depolarized if at least one error has occurred in a qubit that is within a distance of $\min(O(D), n)$ of this qubit. Since there are $\Theta(D)$ steps of the Markov chain, the probability of a certain qubit not being depolarized is about $(1 - p)^{O(D^2)}$ for circuits that are shallow, and about $(1 - p)^{O(nD)}$ for deep circuits. Similarly, in two dimensions a single error propagates on average to $\min(O(D^2), n)$ qubits, and in the nonlocal case to $\min(e^{O(D)}, n)$. These scalings are displayed in Table I. We note that, if the circuit is sufficiently deep, a single error can potentially propagate to all the other qubits, thus resulting in $\langle q \rangle$ scaling as $1 - (1 - p)^{O(nD)}$.

TABLE I. Summary of the scaling of the expectation value of the fraction of depolarized qubits at the end of the computation, as a function of the circuit depth and system size. We identify two different regimes: a shallow-depth regime and a deep regime. The deep regime represents the cases in which the light cone of a single error can reach the edges (i.e., a single error can propagate to all other qubits), while for the shallow regime circuits, this is not the case.

	Architecture	Depth	$\langle q \rangle / n$
Shallow regime	1D	$O(n)$	$1 - (1 - p)^{O(D^2)}$
	2D	$O(\sqrt{n})$	$1 - (1 - p)^{O(D^3)}$
	NL	$O(\log n)$	$1 - (1 - p)^{O[\exp(D)]}$
Deep regime	1D	$\Omega(n)$	$1 - (1 - p)^{O(nD)}$
	2D	$\Omega(\sqrt{n})$	$1 - (1 - p)^{O(nD)}$
	NL	$\Omega(\log n)$	$1 - (1 - p)^{O(nD)}$

In addition to the scalings presented in Table I, we provide a heuristic formula that works very well in practice for the 1D case:

$$\frac{\langle q \rangle_{1D}}{n} \simeq \begin{cases} 1 - (1 - 2p)^{9D^2/80} & \text{if } D \leq \frac{5}{3}n, \\ 1 - (1 - 2p)^{3nD/8 - 5n^2/16} & \text{if } D > \frac{5}{3}n. \end{cases} \quad (12)$$

The derivation of this formula and its numerical verification is given in Appendix C. We also provide a semiempirical formula for the 2D case that has been obtained by fitting the data points to the expression in Table I:

$$\frac{\langle q \rangle_{2D}}{n} \simeq \begin{cases} 1 - (1 - \frac{3}{2}p)^{0.026D^3 + 0.054D^2} & \text{if } D \leq 3.226\sqrt{n}, \\ 1 - (1 - \frac{3}{2}p)^{nD/2 - 0.74n^{3/2} + 0.56n} & \text{if } D > 3.226\sqrt{n}. \end{cases} \quad (13)$$

Finally, in Appendix D we rigorously prove a lower bound on $\langle q \rangle$ for the 1D model with $D < n$ that has the same scaling with p , D , and n as Eq. (12).

B. Implications on circuit depths for noisy QAOA

We apply our error model to the specific case of quantum circuits that try to use QAOA to solve classical optimization problems, and analyze how the propagation of errors could limit the performance of the algorithm. We consider the max-cut problem since it is practically useful and hard to solve classically. Given a graph $G = (V, E)$ on n vertices and adjacency matrix a_{ij} , finding the max-cut of G requires dividing the qubits into two groups, V' and $V \setminus V'$, such that the sum of weights of edges between these two groups is maximal. It can be equivalently defined as the

bitstring $Z \in \{-1, 1\}^n$ maximizing the cost function

$$C = \frac{1}{2} \sum_{(i,j) \in E} a_{ij} (1 - Z_i Z_j) \quad (14)$$

with the vertices corresponding to $Z = -1$ being in one group and those corresponding to $Z = 1$ being in the other.

Let us denote by C_{\max} the solution of this problem, and denote the average value of the cost function when random guessing by C_{avg} , i.e.,

$$C_{\text{avg}} = \text{tr}\left(C \frac{I}{2^n}\right) = \frac{1}{2} \sum_{(i,j) \in E} a_{ij}. \quad (15)$$

The cost function C_{avg} is that obtained by a circuit in which all the qubits are depolarized. Since the output of the error propagation channel contains some qubits that are depolarized, the cost function obtained from the resulting state will be between C_{avg} and C_{\max} . As shown in Appendix A, under the assumption that $a_{ij} > 0$, we are able to upper bound the average energy of the output as a function of the average number of depolarized qubits $\langle q \rangle$ as

$$\begin{aligned} \text{tr}(C\Phi_{\text{avg}}(\rho)) &\leq \frac{1}{2} \left(1 - \frac{\langle q \rangle}{n}\right) \left(2 - \frac{\langle q \rangle}{n}\right) C_{\max} \\ &\quad + \left(1 - \left(1 - \frac{\langle q \rangle}{n}\right)^2\right) C_{\text{avg}}. \end{aligned} \quad (16)$$

This expression readily provides an upper bound on the quality of the solution that we can compute given an error rate and a circuit depth. Given a lower bound on the maximum value C_{\max} of the cost function, it also allows us to upper bound the approximation ratio $[\alpha = \text{Tr}[C\Phi_{\text{avg}}^A(\rho_0)]/C_{\max}]$. This approximation ratio can often inform of the existence of an efficient classical algorithm that obtains a similar solution—for instance, the Goemans-Williamson algorithm is a classical approximation algorithm that has a performance guarantee of $\alpha > 0.878$ [44]. It is thus reasonable to make the assumption that near-term quantum circuits are only useful if the approximation ratio, in the presence of noise, is better than those achievable by classical algorithms.

As a specific example, we briefly study unweighted (i.e., $a_{ij} = 1$) bounded degree- $(\Delta = 3)$ graphs that cannot be solved trivially in general. For any bounded degree- $(\Delta = 3)$ graph, the Edwards-Erdős inequality provides a lower bound for the max-cut in terms of the number of edges, $C_{\max} \geq (2/3)|E|$ [45]. Combined with Eq. (16), this provides an upper bound for the approximation ratio,

$$\alpha \leq 1 - \left(\frac{\langle q \rangle}{2n}\right)^2. \quad (17)$$

For degree-3 graphs, there is a classical approximation algorithm that achieves an approximation ratio of 0.9326

[46]. We therefore find that only when $\langle q \rangle/n \leq 0.52$ can the quantum algorithm possibly output a better average energy than the classical one. That is, as soon as approximately half of the qubits are depolarized, we can be sure that the average quality of the solution is worse than the quality of the solution of classical approximation algorithms in the worst case. We represent an instance of this in Fig. 4 below. We note that this bound is not tight in every case, since we are considering all possible bounded degree- $(\Delta = 3)$ graphs. For example, if we consider only the bipartite ones, $C_{\max} = |E|$ and $C_{\text{avg}} = |E|/2$, which gives an approximation ratio that is bounded by $\alpha \leq 1 - \langle q \rangle/(2n)$, and therefore $\langle q \rangle/n \leq 0.135$. Therefore, in this case, it is already possible to certify the classical advantage when only around 15% of the qubits are depolarized.

Using the scalings from Table I, we find that, for shallow circuits (as defined in the table), after a depth $D = O(\sqrt{1/p})$ for the 1D architecture and $D = O(\sqrt[3]{1/p})$ for the 2D architecture we already have a situation where the quality of the solution is worse than with classical approximation algorithms. This is respectively quadratically and cubically worse with respect to p than the scaling reported in Ref. [36]. This is a consequence of the rapid spreading of errors.

The impact of errors in 2D (and even all-to-all) architectures is much higher than in 1D architectures due to a more rapid propagation of errors. However, this does not necessarily imply that 1D circuits are better for near-term quantum computation since our analysis thus far does not account for the increased connectivity of the quantum circuit. In order to take that into account, in the next subsection we consider a specific case, QAOA for solving a nonlocal problem, i.e., a graph with long-range vertices (a nonplanar graph). This necessarily places us in the deep circuit regime (as defined in Table I) in all cases.

We stress that this is an average-case analysis, which may not apply to particular situations or specific circuits. However, in Appendix B we derive a concentration bound for shallow local circuits,

$$\Pr[|C - \text{tr}(C\Phi_{\text{avg}}(\rho))| \geq \alpha|E|] \leq 2e^{-O(\alpha^2|E|/\Delta^2 D^{2k})}, \quad (18)$$

where Δ is the degree of the graph and k is the dimension of the architecture. Thus, for shallow local circuits, the output of a typical circuit is exponentially close to the average.

1. Required error rate

We also need to analyze the required circuit depth to run QAOA. Most quantum computing architectures have a planar design, similar to the 2D architecture that we are considering. However, the optimization problems that are useful in practice usually consist of nonplanar graphs. For example, planar max-cut can always be solved in polynomial time on classical computers [47]. Since nonplanar

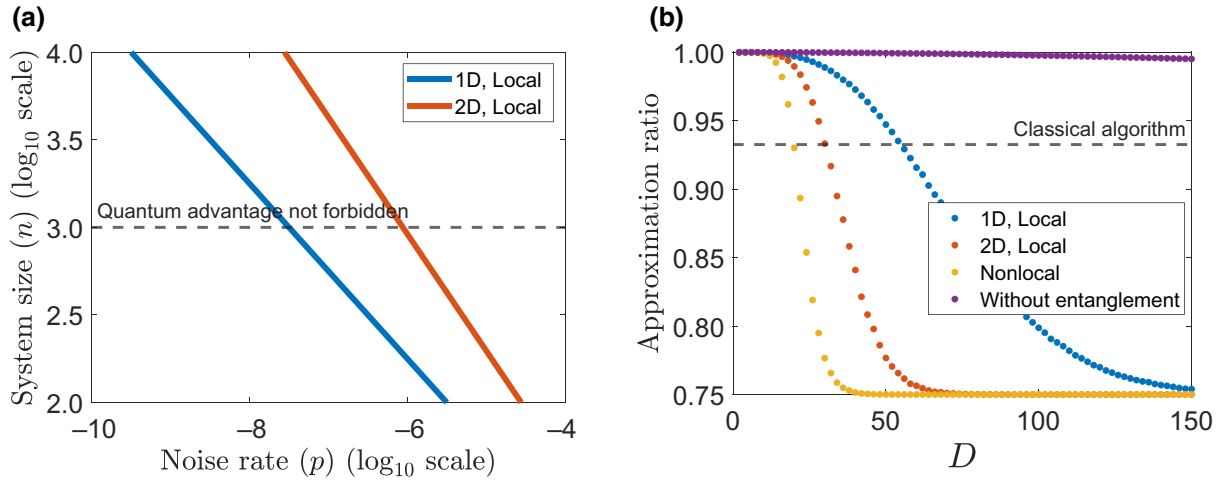


FIG. 4. In (a) we represent the relation between the system size and the error rate when the average number of depolarized qubits at the end of the computation is fixed to $\langle q \rangle/n = 0.5$ for bounded degree- $(\Delta = 3)$ graphs. In the 1D case the depth is $D_{1D} = 30n$, and in the 2D case $D_{2D} = 10\sqrt{7n}$. The horizontal dashed line corresponds to system sizes of $n = 1000$, which is when a potential quantum advantage could begin to be practically useful [26]. In (b) we represent the upper bound for the approximation ratio of bounded degree- $(\Delta = 3)$ graphs, given by Eq. (17), as a function of the circuit depth for the different architectures: 1D, 2D, and nonlocal. We use a system size $n = 900$ and an error rate $p = 10^{-3}$. The horizontal line represents the approximation ratio that is reachable by an efficient classical algorithm. The number of samples taken is 2000, which reduces the error in the estimated mean to 2%.

graphs do not match the connectivity of the hardware, routing will be required to perform the computation. This can be done with SWAP gates that permute the different qubits, but comes at the cost of a growth in the circuit depth. Embedding a bounded degree graph on a square lattice results in an overhead of \sqrt{n} in the gate count, while the cost is n for the 1D local architecture [10]. Furthermore, at least a few QAOA layers are necessary for the algorithm to reach a satisfactory result. Following the scalings reported in Refs. [10,36], we assume that ten layers are enough to reach the solution, and that every QAOA layer, after the routing, needs on average about $\sqrt{7n}$ two-qubit gates in the 2D architecture and $3n$ two-qubit gates in the 1D architecture [48]. In this case, a circuit depth $D_{2D} \geq 10\sqrt{7n}$ for the 2D case is needed and $D_{1D} \geq 30n$ for the 1D case. This places us in the deep circuit regime, as seen in Table I and Eqs. (12) and (13). In this regime, the interpretation is the following: an error in any qubit will, with high likelihood, depolarize all the other qubits. Hence, even if there is just one error in the computation, the average output solution will not be much better than random guessing. Therefore, in order to have a good solution, we need to have a computation completely free of errors. This can only realistically happen if the error rate is as low as $p \sim 1/(nD)$. We perform the exact calculation and represent it in Fig. 4. We see that in this simple example, we would need error rates orders of magnitude below what is currently achievable. Namely of the order of 10^{-7} for the 1D architecture, and 10^{-6} for the 2D architecture. We remark that these figures are obtained under arguably conservative assumptions. A more realistic computation would have to include all the

one-qubit and two-qubit gates, thus restricting the error rate even more.

IV. CONCLUSION

We have studied a model that captures the propagation of errors in noisy quantum devices when the final state is a product state. This is the case, for example, when trying to find the solution to classical optimization problems. For this model, we show that a single error in one qubit is propagated rapidly to the rest of the qubits. This would place stringent restrictions on the error rates that are compatible with a quantum advantage. We estimate the required error rate to be $p \sim 1/(nD)$, where n is the system size and D is the circuit depth. As a consequence, assuming that our error model is representative of the circuits that solve the problem on real hardware, one would expect that noisy devices can only become useful for such problems when the error rates are extremely low, framing fault tolerance as the most realistic solution.

We emphasize that results are obtained by averaging different circuits, and it may be possible that for some particular instances they do not apply. However, due to the concentration result provided in Appendix B for low-depth circuits, we expect the average to be representative of a typical result.

Our results suggest that there is a trade-off between error propagation and entanglement spread. If we want to take advantage of a quantum computer, the quantum circuit should be able to generate entanglement, but this will generally be associated with the propagation of errors.

Equivalently, trying to avoid the propagation of errors may well result in not fully exploiting the whole quantum computer. Perhaps, this deserves a more careful analysis.

Finally, we note that there are other situations where error propagation may not impose the stringent conditions obtained here. For instance, the adiabatic algorithm [6,7] (and its variational extension [49]) is a special kind of circuit where one is always close to the ground state of a particular Hamiltonian. In fact, in this setup there are indications [49] that the propagation of errors is relatively mild. Additionally, in the development of quantum algorithms for quantum problems, like quantum simulation, conservation laws might also prevent the propagation of errors and thus circumvent the restrictions found in the present work. The analysis in this work and these considerations indicate that the propagation of errors should be taken into account in the design of quantum algorithms for NISQ devices.

ACKNOWLEDGMENTS

This research is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project No. 414325145 in the framework of the Austrian Science Fund (FWF): SFB F7104. R.T. acknowledges support from a Max Planck Harvard Research Center for Quantum Optics (MPHQ) Postdoctoral Fellowship.

APPENDIX A: PROOF OF ENERGY BOUND

In this section we outline a proof for the energy bound given in Eq. (16), stated as the following proposition.

Proposition 1 (Energy bound). *Let us consider a weighted graph $G = (V, E)$ with n vertices and adjacency matrix $a_{ij} \geq 0$, and the cost function*

$$C = \frac{1}{2} \sum_{(i,j) \in E} a_{ij} (1 - Z_i Z_j) \quad (\text{A1})$$

with $Z \in \{-1, 1\}^n$. The max-cut problem on this graph is defined as finding the maximum cut of the graph, $C_{\max} = \max_Z C$. Let us also consider the averaged quantum channel Φ_{avg}^A defined in the main text. Then, for any translationally invariant architecture \mathcal{A} , it holds that

$$\begin{aligned} \text{tr}(C \Phi_{\text{avg}}^A(\rho)) &\leq \frac{1}{2} \left(1 - \frac{\langle q \rangle}{n}\right) \left(2 - \frac{\langle q \rangle}{n}\right) C_{\max} \\ &\quad + \left(1 - \left(1 - \frac{\langle q \rangle}{n}\right)^2\right) C_{\text{avg}}, \end{aligned} \quad (\text{A2})$$

where $\langle q \rangle$ is the expectation value of the number of qubits that are depolarized after the application of Φ_{avg}^A , and $C_{\text{avg}} = \text{tr}(C/2^n)$.

In order to show Proposition 1, we first provide a technical lemma that we use for this bound, which provides an inequality on the joint probability of qubits at the output of the circuit to be unaffected by errors. To state this lemma, we first introduce some notation—we consider the Markov chain introduced in Sec. II, but for the purpose of this section, we refer to one application of the transition matrix corresponding to either a unitary gate [Eq. (9)] or a depolarizing channel at a single qubit as a single time step [Eq. (10)]. We denote by $Q_i^t \in \{0, 1\}$ the random variable that indicates the state of the Markov chain at the i th qubit at time step t . It is 0 (i.e., the qubit experiences no noise) or 1 (i.e., the qubit experiences noise). Furthermore, for any subset of qubits S , let $Q_S^t = 0$ be the event $Q_i^t = 0$ for all $i \in S$,

Lemma 1. *Let A, B be two disjoint subsets of qubits; then*

$$\Pr(Q_{A \cup B}^t = 0) \geq \Pr(Q_A^t = 0) \Pr(Q_B^t = 0).$$

Proof. For ease of notation, we introduce

$$V_{A,B}^t = \Pr(Q_{A \cup B}^t = 0) - \Pr(Q_A^t = 0) \Pr(Q_B^t = 0).$$

This measures the violation of the inequality that we intend to show at time step t and in between the qubits in A and B

We note that $V_{A,B}^t \geq 0$ is trivially true at $t = 0$ since all the random variables Q_i^0 are independent of each other. We now assume that this is true at time step t , and show that it is also true for $t + 1$. Suppose first that, from t to $t + 1$, we apply the transition matrix corresponding to the two-qubit unitary on qubits i, j . Then, for any set S such that $i, j \notin S$, we obtain, by an application of the transition matrix in Eq. (9),

$$\begin{aligned} \Pr(Q_{S \cup \{i,j\}}^{t+1} = 0) &= \Pr(Q_i^{t+1} = 0, Q_j^{t+1} = 0, Q_S^t = 0) \\ &= \Pr(Q_{S \cup \{i,j\}}^t = 0) \\ &\quad + \frac{1}{5} \sum_{q \in \{0,1\}} \Pr(Q_i^t = q, Q_j^t = q) \\ &= 1 - q, Q_S^t = 0). \end{aligned}$$

Using the fact that $\Pr(Q_i^t = 0, Q_j^t = 1, Q_S^t = 0) = \Pr(Q_{S \cup \{i\}}^t = 0) - \Pr(Q_{S \cup \{i,j\}}^t = 0)$, we obtain

$$\begin{aligned} \Pr(Q_{S \cup \{i,j\}}^{t+1} = 0) &= \frac{3}{5} \Pr(Q_{S \cup \{i,j\}}^t = 0) \\ &\quad + \frac{1}{5} \sum_{k \in \{i,j\}} \Pr(Q_{S \cup \{k\}}^t = 0). \end{aligned} \quad (\text{A3})$$

We repeatedly use this update rule in the following analysis. There are four cases that we consider.

Case 1. The two-qubit unitary is applied between a qubit in A and a qubit in B . Let us denote by $a \in A$ and $b \in B$ the two qubits between which the unitary is applied, and by $A' = A \setminus \{a\}$ and $B' = B \setminus \{b\}$. From Eq. (A3), it then follows that

$$\Pr(Q_{A \cup B}^{t+1} = 0) = \frac{3}{5} \Pr(Q_{A \cup B}^t = 0) + \frac{1}{5} \sum_{k \in \{a, b\}} \Pr(Q_{A' \cup B' \cup \{k\}}^t = 0).$$

Furthermore, observe from the two-qubit unitary transition matrix in Eq. (9) that $Q_a^t = 0 \implies Q_b^t = 0$ (since both the qubits either simultaneously experience error or not), and, therefore,

$$\Pr(Q_A^{t+1} = 0) = \Pr(Q_{A \cup \{b\}}^{t+1} = 0) = \frac{3}{5} \Pr(Q_{A \cup \{b\}}^t = 0) + \frac{1}{5} \sum_{k \in \{a, b\}} \Pr(Q_{A' \cup \{k\}}^t = 0).$$

Similarly, it holds that

$$\Pr(Q_B^{t+1} = 0) = \frac{3}{5} \Pr(Q_{B \cup \{a\}}^t = 0) + \frac{1}{5} \sum_{k \in \{a, b\}} \Pr(Q_{B' \cup \{k\}}^t = 0).$$

Note that, since $\Pr(Q_{B \cup \{a\}}^t = 0)$, $\Pr(Q_{B' \cup \{a\}}^t = 0)$, $\Pr(Q_{B' \cup \{b\}}^t = 0) \leq \Pr(Q_{B'}^t = 0)$, we obtain $\Pr(Q_B^{t+1} = 0) \leq \Pr(Q_{B'}^{t+1} = 0)$. Therefore,

$$\begin{aligned} V_{A,B}^{t+1} &\geq \Pr(Q_A^{t+1} = 0, Q_B^{t+1} = 0) - \Pr(Q_A^{t+1} = 0) \Pr(Q_{B'}^{t+1} = 0) \\ &= \frac{3}{5} V_{A' \cup \{a, b\}, B'}^t + \frac{1}{5} (V_{A' \cup \{a\}, B'}^t + V_{A' \cup \{b\}, B'}^t), \end{aligned}$$

from which it follows that $V_{A,B}^{t+1} \geq 0$.

Case 2. The two-qubit unitary is applied on two qubits, $a_1, a_2 \in A$. Denote by $A' = A \setminus \{a_1, a_2\}$. Then, it follows from Eq. (A3) that

$$\Pr(Q_{A \cup B}^{t+1} = 0) = \frac{3}{5} \Pr(Q_{A \cup B}^t = 0) + \frac{1}{5} \sum_{k \in \{a_1, a_2\}} \Pr(Q_{A' \cup B \cup \{k\}}^t = 0).$$

Furthermore,

$$\Pr(Q_A^{t+1} = 0) = \frac{3}{5} \Pr(Q_A^t = 0) + \frac{1}{5} \sum_{k \in \{a_1, a_2\}} \Pr(Q_{A' \cup \{k\}}^t = 0),$$

and $\Pr(Q_B^{t+1} = 0) = \Pr(Q_B^t = 0)$. Therefore,

$$V_{A,B}^{t+1} = \frac{3}{5} V_{A,B}^t + \frac{1}{5} (V_{A' \cup \{a_1\}, B}^t + V_{A' \cup \{a_2\}, B}^t),$$

from which it again follows that $V_{A,B}^{t+1} \geq 0$.

Case 3. The two-qubit unitary is applied on a qubit $a \in A$ and a qubit c that is neither in A nor in B . In this case, if $C = A \cup \{c\}$, we simply note that

$$\Pr(Q_{A \cup B}^{t+1} = 0) = \Pr(Q_{C \cup B}^{t+1} = 0) \quad \text{and} \quad \Pr(Q_A^{t+1} = 0) = \Pr(Q_C^{t+1} = 0)$$

since $Q_a^{t+1} = 0 \implies Q_c^{t+1} = 0$. Then $V_{A,B}^{t+1} \geq 0$ simply follows from the analysis in case 2 with the qubit subsets C and B .

Case 4. The two-qubit unitary is applied on qubits that are neither in A nor in B . In this case, it trivially follows that $V_{A,B}^{t+1} = V_{A,B}^t \geq 0$.

Next, we consider the case when the transition matrix corresponding to the depolarizing channel [Eq. (10)] is applied to qubit i . Note that if $i \notin A \cup B$ then $V_{A,B}^{t+1} = V_{A,B}^t \geq 0$ trivially. Consider now the case in which $i \in A$. From Eq. (10), it then follows that

$$\Pr(Q_{A \cup B}^{t+1} = 0) = (1-p)^2 \Pr(Q_{A \cup B}^t = 0) \quad \text{and} \quad \Pr(Q_A^{t+1} = 0) = (1-p)^2 \Pr(Q_A^t = 0),$$

and, therefore, $V_{A \cup B}^{t+1} = (1-p)^2 V_{A \cup B}^t \geq 0$. Hence, we have shown that if at time step t the lemma statement holds then it holds for $t+1$, completing the proof. \blacksquare

We now see how the output energy after applying the averaged quantum channel can be upper bounded as a function of $\langle q \rangle/n$, C_{\max} and C_{avg} . Let us denote by $M \subseteq E$ the subset of edges that are cut in the solution. That is, M contains the edges (i, j) such that $Z_i \neq Z_j$ in the maximum cut. The maximum cut can trivially be written as

$$C_{\max} = \sum_{(i,j) \in M} a_{ij}, \quad (\text{A4})$$

while the average value of the cut when random guessing, C_{avg} , is

$$C_{\text{avg}} = \text{tr}\left(C \frac{I}{2^n}\right) = \frac{1}{2} \sum_{(i,j) \in E} a_{ij}. \quad (\text{A5})$$

While we do not have access to the complete probability distribution, we can upper and lower bound $\Pr(Q'_{i \cup j} = 0)$. Translational invariance implies that $\Pr(Q'_i = 0) = r$ is the same for all qubits, and, therefore,

$$\Pr(Q'_{i \cup j} = 0) = \Pr(Q'_i = 0 \cap Q'_j = 0) \leq \Pr(Q'_i = 0) = r. \quad (\text{A6})$$

Furthermore, from Lemma 1, it follows that

$$\Pr(Q'_{i \cup j} = 0) \geq \Pr(Q'_i = 0)\Pr(Q'_j = 0) = r^2. \quad (\text{A7})$$

Proof of Proposition 1. On the graph, we assume that $a_{ij} \geq 0$ for all (i, j) . This is only for simplicity, but this condition could be dropped. We note that, for a given edge, the contribution to the maximum cut is a_{ij} if $(i, j) \in M$, while if $(i, j) \in E \setminus M$ then it is 0, since it will not be cut. On the other hand, if either i or j is random, after averaging, the contribution is always $a_{ij}/2$, since it will be cut half of the time. Then, the total cost will be

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) = \sum_{(i,j) \in M} a_{ij} \Pr(Q'_{i \cup j} = 0) + \frac{1}{2} \sum_{(i,j) \in E} a_{ij} (1 - \Pr(Q'_{i \cup j} = 0)). \quad (\text{A8})$$

Rearranging the terms this yields

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) = \frac{1}{2} \sum_{(i,j) \in M} a_{ij} (1 + \Pr(Q'_{i \cup j} = 0)) + \frac{1}{2} \sum_{(i,j) \in E \setminus M} a_{ij} (1 - \Pr(Q'_{i \cup j} = 0)). \quad (\text{A9})$$

We can now use Eq. (A7) to obtain

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) \leq \frac{1}{2}(1+r) \sum_{(i,j) \in M} a_{ij} + \frac{1}{2}(1-r^2) \sum_{(i,j) \in E \setminus M} a_{ij} = \frac{1}{2}(1+r)C_{\max} + (1-r^2)\left(C_{\text{avg}} - \frac{C_{\max}}{2}\right). \quad (\text{A10})$$

Then, setting $r = 1 - \langle q \rangle/n$ proves Proposition 1. ■

APPENDIX B: PROOF OF THE CONCENTRATION BOUND

In addition to studying the average energy of the output, it is also interesting to study the variance. Here we provide a concentration bound for local architectures and shallow circuits, for unweighted graphs of bounded degree Δ . We show that, in this case, the energy of a typical circuit is close to the average energy. This is done using Azuma-Hoeffding's inequality [50].

Proposition 2 (Concentration bound). *Let us consider an unweighted graph $G = (V, E)$ of degree Δ . We denote by $C = \text{tr}[C\Phi^A(\rho)]$ the value of the cut after applying one instance of the random quantum channel Φ of depth D , and by $\langle C \rangle = \text{tr}[C\Phi_{\text{avg}}^A(\rho)]$ the value of the cut after applying the averaged quantum channel of depth D . Then, if \mathcal{A} is a local architecture of nearest-neighbor gates in k dimensions,*

$$\Pr[|C - \langle C \rangle| \geq \alpha |E|] \leq e^{-O(\alpha^2 |E| / \Delta^2 D^{2k})}. \quad (\text{B1})$$

Proof. We begin by considering the 1D case, and will then extend the proof to higher dimensions. We are considering a cost function that has $|E|$ terms, where $|E|$ is the number of edges of the graph. We can therefore consider $|E|$ different random variables, $X_i = \text{tr}(\rho C_i)$. Here X_i denotes the contribution of edge i to the value of the cut. The total cost energy will then be a random variable given by $C = \sum_i X_i$. We can now define $Z_t = \mathbb{E}[\sum_i X_i | X_1, \dots, X_t]$. That is, Z_t updates the expectation value of the energy after learning the value of t edges. In order to apply Azuma-Hoeffding's bound, we need to bound the quantity $|Z_t - Z_{t-1}|$, which determines how much the expectation value can change when we learn the energy of one edge. Since we are working with the 1D local model, two qubits can only be correlated if they are closer than $2D$, where D is the depth. This holds if $D < n$. Since the degree is bounded by Δ , learning the energy of one edge could at most update the value of $4D$ qubits, or $2D\Delta$ edges. We can then bound $|Z_t - Z_{t-1}| \leq \Delta D$. Applying Azuma-Hoeffding's inequality yields

$$\Pr[|C - \langle C \rangle| \geq \lambda] \leq 2e^{-\lambda^2/2|E|\Delta^2 D^2}. \quad (\text{B2})$$

If we set λ to be a fraction of the total number of edges, $\lambda = \alpha|E|$, with $\alpha \in (0, 1)$, this yields

$$\Pr[|C - \langle C \rangle| \geq \alpha|E|] \leq 2e^{-\alpha^2|E|/2\Delta^2 D^2}. \quad (\text{B3})$$

The same reasoning can be applied to the 2D case. In two dimensions, assuming that $D < O(\sqrt{n})$, two qubits can only be correlated if they are closer than $O(D)$, where D is the depth. In this case, learning the energy of one edge could at most update the value of $O(D^2)$ qubits, or $O(D^2\Delta)$ edges. We can then bound $|Z_t - Z_{t-1}| \leq O(\Delta D^2)$, and we obtain

$$\Pr[|C - \langle C \rangle| \geq \alpha|E|] \leq 2e^{-O(\alpha^2|E|/\Delta^2 D^4)}. \quad (\text{B4})$$

The same argument holds generally for k dimensions, completing the proof. ■

APPENDIX C: HEURISTIC FORMULA

Here we derive the heuristic formula that provides the average number of depolarized qubits in the 1D case, in Eq. (12) from the main text.

Conjecture 1 (Heuristic 1D formula). *Let us consider the averaged quantum channel with the 1D architecture as defined in the main text, $\Phi_{\text{avg}}^{\text{1D}}$. Then, the expected number of depolarized qubits after applying $\Phi_{\text{avg}}^{\text{1D}}$ is approximately given by*

$$\frac{\langle q \rangle_{\text{1D}}}{n} \simeq \begin{cases} 1 - (1 - 2p)^{9D^2/80} & \text{if } D \leq \frac{5}{3}n, \\ 1 - (1 - 2p)^{3nD/8 - 5n^2/16} & \text{if } D > \frac{5}{3}n. \end{cases} \quad (\text{C1})$$

We now outline the derivation of the formula. The goal is to compute the expected number of depolarized qubits at the end of the computation, denoted as $\langle q \rangle$, using the Markov chain defined in Sec. II. The transition matrix M of the Markov chain is given by $M = \prod_{t=1}^{D_M} M_t$, where M_t is defined as the product of two matrices $M_t = M^{\text{noise}} M_t^U$.

In every step there are therefore two matrices: M^{noise} , which applies the noise with probability $p_M = 2p - p^2$, and M_U , which propagates it. Since we are in the 1D case, we apply layers of unitaries to all the even (or odd) pairs, and M_U will just consist in applying the matrix that propagates the noise [Eq. (9)] to all even (or odd) pairs.

In order to see how this Markov chain behaves, we can consider the case where only one depolarizing error occurs in the computation. In this case, a bit will be flipped from zero to one at a certain time, and we just apply M_U from then on. Since M_U can only propagate this error to neighboring bits, in this case we can only have a string of ones surrounded by zeroes. As a consequence, we can build a Markov chain that counts the number of ones in the state. We call this the ones chain. Using the expression for M_U , it is easy to see that it is a one-dimensional lazy random walk on the line. One can note that this involves applying the matrix that propagates the noise [Eq. (9)] to both edges of the string of ones.

Lemma 2. *The ones chain has transition matrix P on state space $\{0, 1, \dots, n\}$. Matrix P is given by*

$$P(x \rightarrow x') = \begin{cases} \left(\frac{4}{5}\right)^2 & \text{if } x' = x + 2, \\ \left(\frac{1}{5}\right)^2 & \text{if } x' = x - 2, \\ 2\left(\frac{4}{5}\right)\left(\frac{1}{5}\right) & \text{if } x' = x, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{C2})$$

for $x \in (2, n-2)$ and $P(0 \rightarrow 0) = 1$, $P(n \rightarrow n) = 1$, $P(1 \rightarrow 0) = 1/5$, $P(1 \rightarrow 2) = 4/5$, $P(n-1 \rightarrow n) = 4/5$, $P(n-1 \rightarrow n-2) = 1/5$.

We can bound the probability that the walker will be absorbed by the barrier in 0, which will be useful later.

Lemma 3 (Probability of absorption). *The probability of reaching 0 when starting in 1 in the ones chain is upper bounded by 1/4. That is, denoting by X_t the state after t applications of the Markov chain,*

$$\Pr(X_t = 0 | X_0 = 1) \leq \frac{1}{4} \quad \text{for all } t. \quad (\text{C3})$$

Proof. This can be shown using standard techniques for random walks with absorbing barriers [51]. The random walk starts in state 1. In the first step it goes to 0 with probability 1/5, and to 2 with probability 4/5. From that moment on, it can only reach even numbers, so we can

discard all the states consisting of odd numbers. We can also rearrange the absorbing states (0 and n), so that they are first in the transition matrix. The transition matrix is then of the form

$$\begin{pmatrix} I & S \\ 0 & Q \end{pmatrix}.$$

After an infinite time, the probability that the walker is in 0 or in n will be given by $S(I - Q)^{-1}$, where Q is a tridiagonal Toeplitz matrix. Using standard techniques, we find that the probability of reaching 0, starting from 2, with infinite time, is

$$\lim_{t \rightarrow \infty} \Pr(X_t = 0 | X_0 = 2) = (1/5)^2 \times \left[\frac{(16/25)^n - (1/25)^n}{(16/25)^{n+1} - (1/25)^{n+1}} \right] \leq \frac{1}{16}. \quad (\text{C4})$$

Then

$$\lim_{t \rightarrow \infty} \Pr(X_t = 0 | X_0 = 1) \leq \frac{1}{5} + \frac{4}{5} \frac{1}{16} = \frac{1}{4}. \quad (\text{C5})$$

This completes the proof. ■

We now have the ingredients to build the formula. The ones chain can be solved analytically using the techniques above. However, if we ignore the edges, the relation $\langle q(t+1) \rangle = \langle q(t) \rangle + 6/5$ holds. Hence, we can use the very simple formula $\langle q(t) \rangle \simeq (3/4) \min(6/5t, n)$ instead. This is not exact, but it does not deviate much from the exact result. The $3/4$ factor accounts for the fact that, up to a $1/4$ of the time, the walker can be absorbed by the barrier at 0, in which case it stays there. We are neglecting the interaction with the absorbing barrier at n . Therefore, we see that an error propagates, on average, forming a cone. The area of this cone can be computed as

$$\begin{aligned} A &= \int_0^{D_M} \frac{3}{4} \min(6/5t, n) \\ &= H\left(D_M - \frac{5}{6}n\right) \left(\frac{3}{4}nD_M - \frac{5}{16}n^2\right) \\ &\quad + H\left(-D_M + \frac{5}{6}n\right) \left(\frac{9}{20}D_M^2\right), \end{aligned} \quad (\text{C6})$$

where $H(x)$ is the Heaviside step function.

Knowing this, we can see what happens approximately when there are possibly many errors during the computation. We do this by constructing a deterministic model as follows.

Lemma 4. *Let us construct a deterministic model of the propagation of errors as follows: if an error occurs in bit j at time t^* , the value of bit x_k at time t is given by*

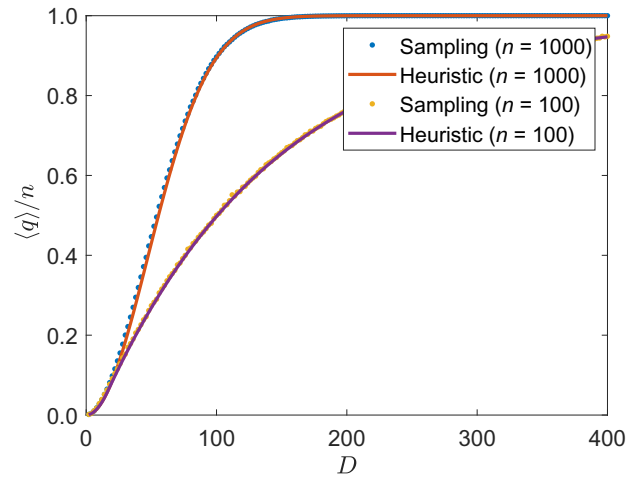


FIG. 5. The heuristic formula in Eq. (12) is plotted along with the results from sampling from the Markov chain for the 1D architecture. This is done for an error rate $p = 10^{-3}$ and with two different system sizes, $n = 100$ and $n = 1000$. The heuristic formula shows good agreement with the result from sampling from the Markov chain. The number of samples taken is 50 000, and the error is upper bounded by 0.4%.

$$x_k = \begin{cases} 1 & \text{if } |j - k| < 3/4 \min(6/5(t - t^*), n), \\ 0 & \text{if } |j - k| > 3/4 \min(6/5(t - t^*), n). \end{cases} \quad (\text{C7})$$

Then, the average number of ones in this model is given by $\langle q(t) \rangle / n = 1 - (1 - p_M)^A$.

In the model above every time there is an error it will propagate as a cone, whose area is given by Eq. (C6). For the cases where there is only one error, this model yields the same value for the average number of ones as the Markov chain. This is by definition, since that single error would propagate as the average cone. If there are many errors, this is no longer the case, since in the Markov chain the different cones are not independent, while in the deterministic model they do not interact and are completely independent. However, this effect is very small. We therefore assume that this model provides a good approximation of the average number of ones in the Markov chain. Setting $p_M = 2p - p^2 \simeq 2p$ and $D_M = D/2$ gives the final expression:

$$\langle q \rangle_{1D} / n \simeq \begin{cases} 1 - (1 - 2p)^{9D^2/80} & \text{if } D \leq \frac{5}{3}n, \\ 1 - (1 - 2p)^{3nD/8 - 5n^2/16} & \text{if } D > \frac{5}{3}n. \end{cases} \quad (\text{C8})$$

We represent this formula in Fig. 5, and verify that it shows very good agreement with the result from sampling directly from the Markov chain.

APPENDIX D: RIGOROUS 1D BOUND

In this section we provide a rigorous bound for $\langle q \rangle/n$ in the 1D case for circuits with depth $D < n$. While this bound is not tight, it shows that the scaling in Table I is the correct one.

Proposition 3 (Rigorous 1D bound). *Let us consider the averaged quantum channel with the 1D architecture as defined in the main text, $\Phi_{\text{avg}}^{\text{1D}}$. Then, the expected number of depolarized qubits after applying $\Phi_{\text{avg}}^{\text{1D}}$ is lower bounded by*

$$\langle q \rangle_{\text{1D}}/n \geq c(1 - (1 - p)^{O(D^2)}) \quad (\text{D1})$$

for some $c < 3/4$.

To prove this proposition, we first introduce some notation. We are considering the Markov chain introduced in Sec. II. In this picture, an error in qubit k corresponds to flipping bit k from 0 to 1. We define the set of all possible errors by $\mathcal{S} = \{1, \dots, n\} \times \{1, \dots, D_M\}$. Then, an error is given by the 2-tuple $(a, b) \in \mathcal{S}$. In this notation a specifies the bit where the error occurred, and b specifies the time step. We define an instance of errors by a subset $s \subset \mathcal{S}$ that contains all the errors that have occurred in a given run of the Markov chain. We denote by $\Pr(Q_j^t = 1)$ the probability that bit j is in 1 at the end of the computation. This is given by

$$\Pr(Q_j^t = 1) = \sum_{s \subset \mathcal{S}} \Pr(Q_j^t = 1|s) \Pr(s). \quad (\text{D2})$$

We note that, with this Markov chain, for every instance of errors $s \subset \mathcal{S}$, it holds that $\Pr(Q_j^t = 1|s) \geq \Pr(Q_j^t = 1|(a, b) \in s)$. That is, given an instance of errors s , the probability of bit j being in 1 can only decrease if we pick only one of the errors in s .

Let us now assume that there exists a subset of errors $A \subset \mathcal{S}$ such that $\Pr(Q_j^t = 1|(a, b) \in A) \geq c$, where c is a constant. That is, A contains errors such that, if one of the errors in A occurs, it is enough to certify that the probability of bit j being in 1 is equal or greater than c . Then, we can compute

$$\begin{aligned} & \sum_s \Pr(Q_j^t = 1|s) \Pr(s) \\ & \geq \sum_s \Pr(Q_j^t = 1|(a, b) \in s) \Pr(s) \\ & \geq \sum_s \Pr(s \cap A \neq \emptyset) c \\ & = c \left(\sum_s 1 - \Pr(s \cap A = \emptyset) \right) \\ & = c[1 - (1 - p_M)^{|A|}]. \end{aligned} \quad (\text{D3})$$

We would now like to identify a subset of errors A that fulfills this property. To do this, we analyze the behavior of the Markov chain. We would like to compute $\Pr[Q_j^t = 1|(k, t^*)]$. As explained in Appendix C, if there is only one error, it will propagate, forming a string of ones. To compute the probability, we can track the movement of the endpoints of this string. We define several random variables for this purpose.

Definition 1 (Random walks): Let us consider the random variables

$$X_A(t) = X_A^0 + \sum_{i=1}^{t^*} X_A^{(i)}, \quad (\text{D4})$$

$$X_B(t) = X_B^0 + \sum_{i=1}^{t^*} X_B^{(i)}, \quad (\text{D5})$$

where $X_A^0 = k$ and

$$X_B^0 = \begin{cases} k+1 & \text{with probability } 4/5, \\ k-1 & \text{with probability } 1/5, \end{cases} \quad (\text{D6})$$

$$X_A^{(i)} = \begin{cases} 0 & \text{if } X_A(t-1) > X_B(t-1), \\ 1 & \text{with probability } 1/5 \text{ if } X_A(t-1) < X_B(t-1), \\ -1 & \text{with probability } 4/5 \text{ if } X_A(t-1) < X_B(t-1), \end{cases} \quad (\text{D7})$$

$$X_B^{(i)} = \begin{cases} 0 & \text{if } X_A(t-1) > X_B(t-1), \\ -1 & \text{with probability } 1/5 \text{ if } X_A(t-1) < X_B(t-1), \\ 1 & \text{with probability } 4/5 \text{ if } X_A(t-1) < X_B(t-1). \end{cases} \quad (\text{D8})$$

We also define two independent random walkers as

$$X_A^{\text{ind}}(t) = X_A^0 + \sum_{i=1}^{t^*} X_A^{\text{ind},(i)}, \quad (\text{D9})$$

$$X_B^{\text{ind}}(t) = X_B^0 + \sum_{i=1}^{t^*} X_B^{\text{ind},(i)}, \quad (\text{D10})$$

where

$$X_A^{\text{ind},(i)} = \begin{cases} 1 & \text{with probability } 1/5, \\ -1 & \text{with probability } 4/5, \end{cases} \quad (\text{D11})$$

$$X_B^{\text{ind},(i)} = \begin{cases} -1 & \text{with probability } 1/5, \\ 1 & \text{with probability } 4/5. \end{cases} \quad (\text{D12})$$

With the definitions above, it holds that

$$\Pr[Q_j^t = 1 | (k, t^*)] = \Pr(X_B(t) \geq j \cap X_A(t) \leq j). \quad (\text{D13})$$

The random walks X_A and X_B defined above track the end points of the string of ones. The difficulty here is that they are not independent random walks. They behave independently until they cross: in that case they stop. However, we can bound the probability using X_A^{ind} and X_B^{ind} , which are just biased random walks in the line, with every step being independent of the rest.

Lemma 5. *Let us consider the random variables X_A , X_A^{ind} , X_B , and X_B^{ind} as defined in Definition 1. Then,*

$$\begin{aligned} \Pr(Q_j^t = 1 | (k, t^*)) &= \Pr(X_B(t) \geq j \cap X_A(t) \leq j) \\ &\geq \Pr(X_B(t)^{\text{ind}} \geq j) \Pr(X_A(t)^{\text{ind}} \leq j) - \frac{1}{4}. \end{aligned} \quad (\text{D14})$$

Proof. We want to compute $\Pr(X_B(t) \geq j \cap X_A(t) \leq j)$. We note that the two random walks are completely independent until the moment they cross: in that case they stop. We can denote this event as C , and no crossing as NC . Assuming that $t < n/2$, from Lemma 3 we know that the probability that they cross is upper bounded, $\Pr(C) \leq 1/4$. Then, we can write:

$$\begin{aligned} \Pr(X_A(t) \geq j \cap X_B(t) \leq j) &= \Pr(X_A(t) \geq j \cap X_B(t) \leq j | NC) \Pr(NC). \end{aligned} \quad (\text{D15})$$

We can now consider the two random walks X_A^{ind} and X_B^{ind} . Their behaviors are the same as X_A and X_B provided there

is no crossing. That is,

$$\begin{aligned} \Pr(X_A(t) \geq j \cap X_B(t) \leq j | NC) &= \Pr(X_A^{\text{ind}}(t) \geq j \cap X_B^{\text{ind}}(t) \leq j | NC). \end{aligned} \quad (\text{D16})$$

Using now the law of total probability, we get

$$\begin{aligned} \Pr(X_A^{\text{ind}}(t) \geq j \cap X_B^{\text{ind}}(t) \leq j | NC) \Pr(NC) &= \Pr(X_A^{\text{ind}}(t) \geq j) \Pr(X_B^{\text{ind}}(t) \leq j) \\ &\quad - \Pr(X_A^{\text{ind}}(t) \geq j \cap X_B^{\text{ind}}(t) \leq j | C) \Pr(C) \\ &\geq \Pr(X_A^{\text{ind}}(t) \geq j) \Pr(X_B^{\text{ind}}(t) \leq j) - \frac{1}{4}. \end{aligned} \quad (\text{D17})$$

The result then follows immediately from Eq. (D13). ■

We know now how to bound the probability $\Pr[Q_j^t = 1 | (k, t^*)]$ using two biased random walks on the line. This allows us to obtain the combinations of (k, t^*) that we are interested in.

Lemma 6. *Let us consider an error (k, t^*) such that*

$$|j - k| \leq \frac{3}{5}(t - t^*) - \sqrt{2(t - t^*) \ln \frac{1}{1 - \sqrt{1/4 + c}}}. \quad (\text{D18})$$

Then, $\Pr[Q_j^t = 1 | (k, t^*)] \geq c$.

Proof. Since X_A^{ind} and X_B^{ind} are a sum of independent random variables, we can bound the probability that they deviate from the mean using Hoeffding's inequality [52]:

$$\Pr(X_A^{\text{ind}}(t) - E[X_A^{\text{ind}}(t)] \geq \alpha) \leq \exp\left(-\frac{\alpha^2}{2(t - t^*)}\right). \quad (\text{D19})$$

The expectation values can easily be computed as $E[X_A^{\text{ind}}(t)] = k - 3/5(t - t^*)$ and $E[X_B^{\text{ind}}(t)] = k + 3/5(1 + t - t^*)$. Let us denote $d_A = j - E[X_A^{\text{ind}}(t)]$. Then,

$$\begin{aligned} \Pr(X_A^{\text{ind}}(t) \leq j) &= 1 - \Pr(X_A^{\text{ind}}(t) \geq j) \\ &= 1 - \Pr(X_A^{\text{ind}}(t) - E[X_A^{\text{ind}}(t)] \geq d_A) \\ &\geq 1 - \exp\left(-\frac{d_A^2}{2(t - t^*)}\right). \end{aligned} \quad (\text{D20})$$

Let us now assume, without loss of generality, that $j < k$. Then,

$$\Pr(X_B^{\text{ind}}(t) \geq j) \geq \Pr(X_A^{\text{ind}}(t) \leq j). \quad (\text{D21})$$

Therefore, we obtain

$$\Pr[Q_j^t = 1 | (k, t^*)] \geq \left(1 - \exp\left(-\frac{d_A^2}{2(t-t^*)}\right)\right)^2 - \frac{1}{4} \geq c. \quad (\text{D22})$$

Solving for d_A then yields

$$d_A \geq \sqrt{2(t-t^*) \ln \frac{1}{1-\sqrt{1/4+c}}}, \quad (\text{D23})$$

and, therefore,

$$|k-j| \leq \frac{3}{5}(t-t^*) - \sqrt{2(t-t^*) \ln \frac{1}{1-\sqrt{1/4+c}}}. \quad (\text{D24})$$

For symmetry reasons, this calculation holds as well when $j > k$. ■

We have therefore found a region such that $\Pr[Q_j^t = 1 | (k, t^*)] \geq c$. Integration over t^* will yield the area of such a region:

$$|A| = \frac{3}{5}t^2 - \frac{4}{3}\sqrt{2 \ln \frac{1}{3/4-c}}t^{3/2}. \quad (\text{D25})$$

Then, using Eq. (D3), we can bound

$$\langle q \rangle / n \geq c(1 - (1 - p_M)^{3t^{2/5-(4/3)}\sqrt{2 \ln[1/(1-\sqrt{1/4+c})]}t^{3/2}}). \quad (\text{D26})$$

Then, setting $p_M = 2p - p^2 \simeq 2p$ and $t = D/2$ yields the result in Proposition 3.

[1] Frank Arute, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
[2] Yulin Wu, *et al.*, Strong Quantum Computational Advantage Using a Superconducting Quantum Processor, *Phys. Rev. Lett.* **127**, 180501 (2021).
[3] Han-Sen Zhong, *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).
[4] John Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
[5] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).

[6] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser, Quantum computation by adiabatic evolution. arXiv preprint (2000), [ArXiv:quant-ph/0001106](https://arxiv.org/abs/quant-ph/0001106).
[7] Tameem Albash and Daniel A. Lidar, Adiabatic quantum computation, *Rev. Mod. Phys.* **90**, 015002 (2018).
[8] Google AI Quantum, Collaborators*†, Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, *et al.*, Hartree-fock on a superconducting qubit quantum computer, *Science*, **369**, 1084 (2020).
[9] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, and Lukasz Cincio, *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
[10] Matthew P. Harrigan, Kevin J. Sung, Matthew Neeley, Kevin J. Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C. Bardin, Rami Barends, and Sergio Boixo, *et al.*, Quantum approximate optimization of non-planar graph problems on a planar superconducting processor, *Nat. Phys.* **17**, 332 (2021).
[11] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New J. Phys.* **18**, 023023 (2016).
[12] Tyson Jones, Suguru Endo, Sam McArdle, Xiao Yuan, and Simon C. Benjamin, Variational quantum algorithms for discovering Hamiltonian spectra, *Phys. Rev. A* **99**, 062304 (2019).
[13] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 1 (2014).
[14] Dave Wecker, Matthew B. Hastings, and Matthias Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
[15] David Amaro, Carlo Modica, Matthias Rosenkranz, Mattia Fiorentini, Marcello Benedetti, and Michael Lubasch, Filtering variational quantum algorithms for combinatorial optimization, *Quantum Sci. Technol.* **7**, 015021 (2022).
[16] Tadashi Kadowaki and Hidetoshi Nishimori, Quantum annealing in the transverse Ising model, *Phys. Rev. E* **58**, 5355 (1998).
[17] Suguru Endo, Iori Kurata, and Yuya O. Nakagawa, Calculation of the Green's function on near-term quantum computers, *Phys. Rev. Res.* **2**, 033281 (2020).
[18] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Quantum annealing: A new method for minimizing multidimensional functions, *Chem. Phys. Lett.* **219**, 343 (1994).
[19] Hongbin Liu, Guang Hao Low, Damian S. Steiger, Thomas Häner, Markus Reiher, and Matthias Troyer, Prospects of quantum computing for molecular sciences, *Mater. Theory* **6**, 1 (2022).
[20] Andrew A. Houck, Hakan E. Türeci, and Jens Koch, On-chip quantum simulation with superconducting circuits, *Nat. Phys.* **8**, 292 (2012).
[21] Matteo Ippoliti, Kostyantyn Kechedzhi, Roderich Moessner, S. L. Sondhi, and Vedika Khemani, Many-Body

- Physics in the NISQ Era: Quantum Programming a Discrete Time Crystal, *PRX Quantum* **2**, 030346 (2021).
- [22] Bernhard H. Korte, Jens Vygen, B. Korte, and J. Vygen, *Combinatorial Optimization* (Springer, Heidelberg, 2011), Vol. 1.
- [23] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, A quantum approximate optimization algorithm. (2014), arXiv preprint [ArXiv:1411.4028](#).
- [24] Edward Farhi and Aram W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm. (2016), arXiv preprint [ArXiv:1602.07674](#).
- [25] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou, The quantum approximate optimization algorithm and the Sherrington-Kirkpatrick model at infinite size. (2019), arXiv preprint [ArXiv:1910.08187](#).
- [26] Gian Giacomo Guerreschi and Anne Y. Matsuura, Qaoa for max-cut requires hundreds of qubits for quantum speed-up, *Sci. Rep.* **9**, 1 (2019).
- [27] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D. Lukin, Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices, *Phys. Rev. X* **10**, 021067 (2020).
- [28] Guido Pagano, Aniruddha Bapat, Patrick Becker, Katherine S. Collins, Arinjoy De, Paul W. Hess, Harvey B. Kaplan, Antonis Kyprianidis, Wen Lin Tan, Christopher Baldwin, Lucas T. Brady, Abhinav Deshpande, Fangli Liu, Stephen Jordan, Alexey V. Gorshkov, and Christopher Monroe, Quantum approximate optimization of the long-range Ising model with a trapped-ion quantum simulator, *Proc. Nat. Acad. Sci.* **117**, 25396 (2020).
- [29] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1 (2021).
- [30] Tobias Haug, Kishor Bharti, and M. S. Kim, Capacity and Quantum Geometry of Parametrized Quantum Circuits, *PRX Quantum* **2**, 040309 (2021).
- [31] Kouhei Nakaji and Naoki Yamamoto, Expressibility of the alternating layered ansatz for quantum computation, *Quantum* **5**, 434 (2021).
- [32] V. Akshay, H. Philathong, M. E. S. Morales, and J. D. Biamonte, Reachability Deficits in Quantum Approximate Optimization, *Phys. Rev. Lett.* **124**, 090504 (2020).
- [33] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles, Noise-induced barren plateaus in variational quantum algorithms, *Nat. Commun.* **12**, 1 (2021).
- [34] Jeffrey Marshall, Filip Wudarski, Stuart Hadfield, and Tad Hogg, Characterizing local noise in QAOA circuits, *IOP SciNotes* **1**, 025208 (2020).
- [35] Cheng Xue, Zhao-Yun Chen, Yu-Chun Wu, and Guo-Ping Guo, Effects of quantum noise on quantum approximate optimization algorithm, *Chin. Phys. Lett.* **38**, 030302 (2021).
- [36] Daniel Stilek França and Raul Garcia-Patron, Limitations of optimization algorithms on noisy quantum devices, *Nat. Phys.* **17**, 1221 (2021).
- [37] Dorit Aharonov, Michael Ben-Or, Russell Impagliazzo, and Noam Nisan, Limitations of noisy reversible computation. (1996), arXiv preprint [ArXiv:quant-ph/9611028](#).
- [38] Michael Ben-Or, Daniel Gottesman, and Avinatan Hassidim, Quantum refrigerator. (2013), arXiv preprint [ArXiv:1301.1995](#).
- [39] Adam Boulund, Bill Fefferman, Zeph Landau, and Yunchao Liu, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, p. 1308. IEEE, 2022.
- [40] Abhinav Deshpande, Bill Fefferman, Alexey V. Gorshkov, Michael J. Gullans, Pradeep Niroula, and Oles Shtanko, Tight bounds on the convergence of noisy random circuits to uniform. (2021), arXiv preprint [ArXiv:2112.00716](#).
- [41] Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão, Random quantum circuits transform local noise into global white noise. (2021), arXiv preprint [ArXiv:2111.14907](#).
- [42] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven, Characterizing quantum supremacy in near-term devices, *Nat. Phys.* **14**, 595 (2018).
- [43] Joseph Emerson, Robert Alicki, and Karol Życzkowski, Scalable noise estimation with random unitary operators, *J. Opt. B: Quantum Semiclass. Opt.* **7**, S347 (2005).
- [44] Michel X. Goemans and David P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. ACM* **42**, 1115 (1995).
- [45] Paul Erdős, On some extremal problems in graph theory, *Isr. J. Math.* **3**, 113 (1965).
- [46] Eran Halperin, Dror Livnat, and Uri Zwick, Max cut in cubic graphs, *J. Alg.* **53**, 169 (2004).
- [47] F. Hadlock, Finding a maximum cut of a planar graph in polynomial time, *SIAM J. Comput.* **4**, 221 (1975).
- [48] It should be noted that a noiseless SWAP gate cannot propagate an error from one qubit to another, since it will just swap the error. Therefore, a more sophisticated model could account for this by distinguishing between the SWAP gates and other two-qubit gates, as well as adding other sources and types of noise. We still expect our scalings to hold in that case.
- [49] Benjamin F. Schiffer, Jordi Tura, and J. Ignacio Cirac, Adiabatic spectroscopy and a variational quantum adiabatic algorithm. (2021), arXiv preprint [ArXiv:2103.01226](#).
- [50] Kazuoki Azuma, Weighted sums of certain dependent random variables, *Tohoku Math. J.* **19**, 357 (1967).
- [51] Gregory F. Lawler, *Introduction to Stochastic Processes* (Chapman and Hall/CRC, New York, 2018).
- [52] Wassily Hoeffding, Probability inequalities for sums of bounded random variables, *J. Am. Stat. Assoc.* **58**, 13 (1963).