

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}

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

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

(Dated: March 30, 2022)

We propose a random circuit model to analyze the impact of noise on the performance of variational quantum circuits for 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 classical optima is low on average and 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 QAOA for classical optimization problems with 2D circuits.

I. INTRODUCTION

Significant advances in the capabilities of quantum information processing hardware have been made recently, 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 [5–20], 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 [21]. 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 [8–14], most notably the quantum approximate optimization algorithm (QAOA)[22]. While variational algorithms in general lack provable guarantees for quantum speedups, there is a growing body of literature that suggests its use for practically interesting problems that remain hard to solve on classical computers [23–27]. However, a number of algorithmic challenges to these heuristics which stem from the limitations on the accessible circuit architectures, such as barren plateaus [28], expressibility of the ansatz [29, 30], or reachability deficits [31] have been identified.

The influence of noise in the quantum devices on variational algorithms is an important consideration for assessing their utility in the near term and has not received as much attention. Numerical modelling of the impact of noise has been attempted [32, 33], but this analysis is limited to very small circuit sizes. An alternative approach is to use entropic arguments to rigorously analyze the impact of noise [34–36]. By simply tracking the von Neumann entropy of the quantum state, in Refs. [35, 36] 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. Ref. [34] improved on this result and showed that even beyond circuit depths of $\Theta(1)$, the quality of solution of a large class of classical optimization algorithm 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 consequently they also bound performance of circuits which 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 to analyze the impact of noise on the performance of variational quantum circuits for classical optimization problems. Our random circuit model accounts for the fact that the circuits map product states for product states. 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. Our study here is thus different from recent works that have studied the impact of noise in Haar-random quantum circuits [37–40] with geometric constraints, where a typical member of these circuit families has an anti-concentrated output state, which is in stark contrast to what is expected for variational quantum circuits that ideally output a state close to a product state.

We perform an average case analysis on this distribution of circuits, and find that noise propagates very rapidly through a typical member of this distribution, severely limiting the performance of the circuits. We provide both numerical results and analytical scalings for three different architectures: 1D, 2D, and nonlocal. Our results indicate that to obtain a solution to a clas-

sical 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 1D and $\max(O(p^{-1/3}), O(1/(pn)))$ in 2D. Furthermore, it follows from our analysis that the impact of error is qualitatively similar for different noise models in contrast to what is predicted by a worst case analysis.

II. ERROR PROPAGATION MODEL

Consider the problem of solving a classical optimization problem over $\{0, 1\}^n$, and suppose that the unitary circuit which 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 the impact of errors, we assume that the optimal circuit is already known, 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

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

where \mathcal{N} is a layer of the noise channel, and $\mathcal{U}^{(t)}$ is a layer of 2-qubit Haar random unitaries. The effective channel for the circuit under consideration, depicted in Fig. 1a, 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 1-qubit gates are noiseless.

We will consider three different architectures for generating $\mathcal{U}^{(t)}$ which model different interaction ranges that can be accessed on physical hardware:

- A 1D local architecture with periodic boundary conditions, where alternating layers of nearest neighbour gates are applied, which we will call 1D.
- A 2D local architecture with periodic boundary conditions on a square lattice, which we will call 2D. Alternating layers of horizontal and vertical 2-qubit gates are applied.
- A nonlocal architecture where, for each layer, $n/2$ pairs are chosen at random, and $n/2$ 2-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 will con-

sider the output of the channel averaged over the random unitaries, for each of the architectures,

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

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. 1b).

We next compute the channel $\Phi_{\text{avg}}^{\mathcal{A}}$ — we will use the fact that the twirl of a 2-qubit quantum channel over the Haar measure is a 2-qubit depolarizing channel (Fig. 2a) [41] i.e.,

$$\begin{aligned} \mathcal{E}_{\text{dep}}(\rho) &= \int_{U(4)} \mu_{\text{Haar}}(dU) U^\dagger \mathcal{N}^{\otimes 2}(U\rho U^\dagger) U, \\ &= \lambda\rho + (1-\lambda)\frac{I}{4}, \end{aligned} \quad (3)$$

where if A_k represents the k -th Kraus operator of the channel $\mathcal{N}^{\otimes 2}$, then

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

When \mathcal{N} is depolarizing noise applied with probability p , $\mathcal{N}(\rho) = (1-p)\rho + p \text{tr}(\rho)/2$, the effective depolarizing channel \mathcal{E}_{dep} can be constructed by considering three different cases:

- With probability $(1-p)^2$, no errors occur in either qubit, consequently $\mathcal{E}_{\text{dep}}(\rho)$ trivially maps ρ to ρ .
- With probability p^2 errors occur in both qubits, $\mathcal{E}_{\text{dep}}(\rho)$ trivially maps ρ to $I/4$.
- With probability $2p(1-p)$, an error occurs only in one of the qubits, the channel maps ρ to $(1/5)\rho + (4/5)I/4$. This can be interpreted as the error being propagated to the other qubit with probability $4/5$ due to the two-qubit entangling gates.

Let us now consider the quantum channel $\Phi_{\text{avg}}^{\mathcal{A}}$ — the key idea to computing this average channel, depicted schematically in Fig. 2b, 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 $\Phi_t^{\mathcal{A}}$ the channel formed by including and averaging over t layers of the entangling and uncomputing unitary from the center of the circuit — since from Eq. (3), averaging over the unitary yields a depolarizing channel, this channel either traces out a qubit and replaces it with $I/2$, or leaves it unchanged i.e.

$$\Phi_t^{\mathcal{A}} = \sum_{\vec{j} \in \{0,1\}^n} p_t^{\mathcal{A}}(\vec{j}) \left[\bigotimes_{\alpha \in \vec{j}} \tau_\alpha \right] \quad (5)$$

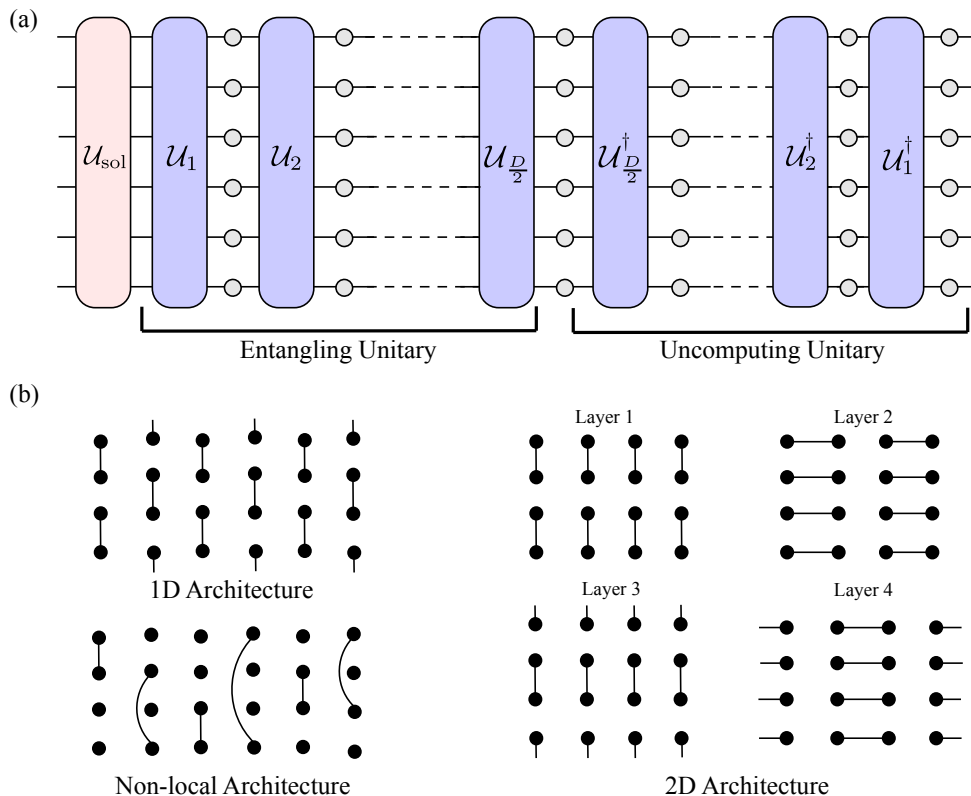


FIG. 1. (a) Schematic depiction of the ensemble of unitary circuits considered in this paper. The circuits consist of an entangling and uncomputing unitary which maps a product state to another product state in the absence of noise. (b) Different circuit architectures (1D, 2D and non-local circuits) studied in this paper.

for some probability distribution p_t^A over $\{0, 1\}^n$ and

$$\tau_\alpha = \begin{cases} \text{id}(\cdot) & \text{if } \alpha = 0, \\ \text{tr}(\cdot) \frac{I}{2} & \text{if } \alpha = 1. \end{cases} \quad (6)$$

Now, by explicitly using Eq. (3) and accounting for the architectures of the unitary layers, the probability distribution describing the channel Φ_{t+1}^A, p_{t+1}^A , can be related via a stochastic matrix M_t to p_t^A . In particular, by noting that the channel Φ_{t+1}^A is obtained by an additional unitary and an additional depolarizing layer, M_t can be written as

$$M_t = M_t^{\text{noise}} M_t^U. \quad (7)$$

From Eq. (3), it follows that M_t^U can be constructed by applying a 4×4 stochastic matrix P , given by

$$P(x \rightarrow y) = \begin{cases} 1 & \text{if } (x, y) = (00, 00), \\ 1/5 & \text{if } (x, y) \in \{(01, 00), (10, 00)\}, \\ 4/5 & \text{if } (x, y) \in \{(01, 11), (10, 11)\}, \\ 1 & \text{if } (x, y) = (11, 11), \end{cases}$$

to the bits corresponding to qubit pairs which have a two-qubit gate being applied on them at layer t (away from

the center of the circuit) in the architecture \mathcal{A} . M_t^{noise} flips from 0 to 1 with probability $p_M(t)$ each bit of the string independently, where:

$$p_M(t) = \begin{cases} p & \text{if } t \in \{1, D/2\} \\ 2p - p^2 & \text{if } t \notin \{1, D/2\}. \end{cases} \quad (8)$$

If a bit is already in 1, then it stays in 1 after applying M_t^{noise} . We note that, per the definition above, a quantum channel of depth D and noise strength p is mapped to a Markov chain with $D/2$ steps and noise strength p_M , where $p_M = p$ in the first and last steps, and $p_M = 2p - p^2$ in the rest. To avoid confusion, we refer to the number of applications of the Markov chain as D_M , and to the noise strength of the Markov chain as p_M . It is also a good approximation to take $p_M \simeq 2p$ in every step of the Markov chain.

To analyze the impact of noise on the output of the quantum circuit, we thus need only to compute the probabilities $p_{D_M=D/2}^A$ by simulating the Markov chain with transition matrices described above, which 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 using Markov chain Monte Carlo

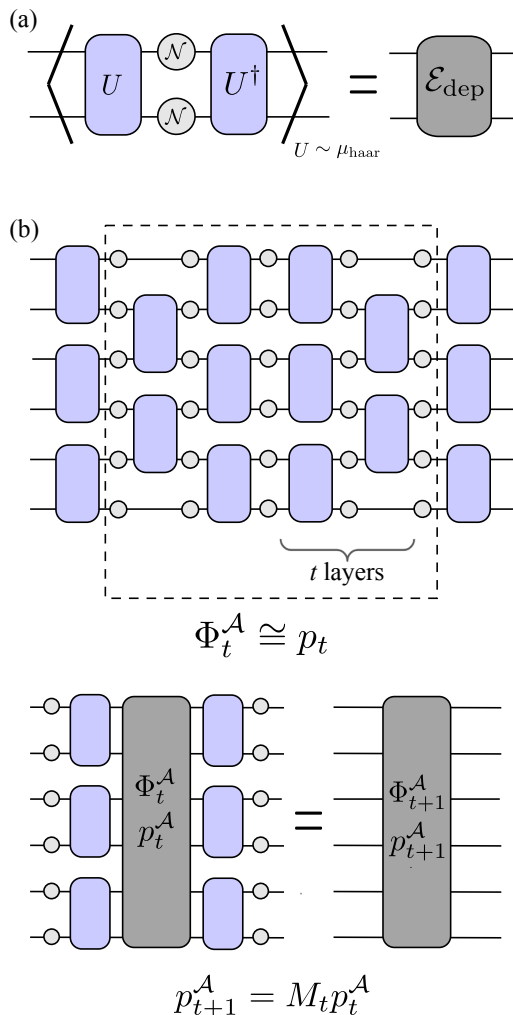


FIG. 2. (a) Averaging over Haar-random unitary U yields a 2-qubit depolarizing channel \mathcal{E}_{dep} . (b) Schematic depiction of averaging over the circuit ensemble shown in Fig. 1 — we average from the center of the circuit to outwards, and derive an effective Markov chain that describes the channels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

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 are now ready to analyze the results of the model. We are interested in computing 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. The numerical results after sampling can be seen in Fig. 3. Our results show that, as expected, the convergence to uniform is very fast with the propagation model that we are considering.

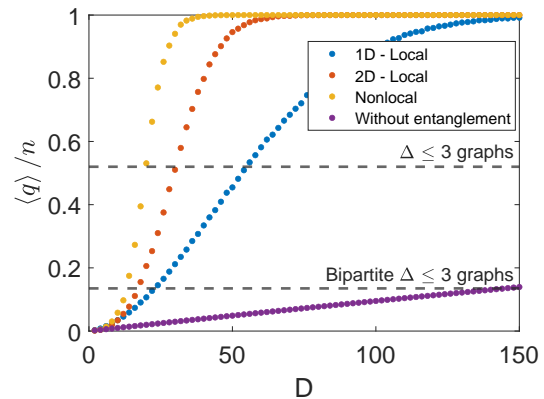


FIG. 3. We represent the expectation value of the fraction of depolarized qubits, $\langle q \rangle / n$, as a function of the circuit depth, for the different architectures: 1D, 2D, and nonlocal. We have used a system size $n = 900$ and an error rate $p = 10^{-3}$. Additionally, we have represented $\langle q \rangle / n = 1 - (1 - p)^D$, 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 subsection III B).

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

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.

In order to better understand these numerical results, we would like to analyze theoretically the behavior of the Markov chain. The scalings of $\langle q \rangle$ with n can be estimated by noting that every time an error occurs, it can propagate to a neighbouring qubit with a certain probability that is $\Theta(1)$. As a consequence, in the 1D model, after a circuit depth D , the error will have propagated to $\min(O(D), n)$ qubits. We can consider the probability that a certain qubit is depolarized at the end of the computation. On average, it will be depolarized if at least an error has occurred in a qubit that is $\min(O(D), n)$ close. Since there are $D/2$ steps, this means that the probability of a certain qubit not being depolarized is

of order $(1-p)^{O(D^2)}$ for circuits that are shallow, and $(1-p)^{O(nD)}$ for deep circuits. Hence, the formulas above are obtained. Analogously, in 2D a single error propagates on average to $\min(O(D^2), n)$ qubits, and in the nonlocal case to $\min(O(e^D), n)$. We note that, if the circuit is sufficiently deep, a single error can potentially propagate to all the other qubits, hence the scaling as $1 - (1-p)^{O(nD)}$.

In addition to the scalings presented in Table I, we provide a heuristic formula that works very well in practice for the 1D case. The derivation of this formula and its numerical verification in appendix B.

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

We also provide a semi-empirical formula for the 2D case, that has been obtained by fitting the data points to the expression in Table I. It is as follows:

$$\frac{\langle q \rangle_{2D}}{n} \simeq \begin{cases} 1 - (1 - \frac{3}{2}p)^{0.026D^3 + 0.054D^2} & \text{if } D \leq 3.226\sqrt{n} \\ 1 - (1 - \frac{3}{2}p)^{\frac{1}{2}nD - 0.74n^{3/2} + 0.56n} & \text{if } D > 3.226\sqrt{n}. \end{cases} \quad (10)$$

Finally, in appendix C we rigorously prove a lower bound

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) \leq \frac{1}{2} \left(1 - \frac{\langle q \rangle}{n}\right) \left(2 - \frac{\langle q \rangle}{n}\right) C_{\text{max}} + \left(1 - \left(1 - \frac{\langle q \rangle}{n}\right)^2\right) C_{\text{avg}}. \quad (13)$$

Furthermore, in appendix D, we show that for low depth circuits, $\text{tr}(C\Phi(\rho))$ for a randomly drawn circuit instance concentrates around this mean indicating that it is representative of a typical circuit instance.

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 ($\alpha = \text{Tr}[C\Phi_{\text{avg}}^A(\rho_0)]/C_{\text{max}}$). This approximation ratio can often inform of the existence of an efficient classical algorithm which obtains a similar solution — for instance, there is a classical approximation algorithm that has a performance guarantee of $\alpha > 0.878$ [42]. Furthermore, most classical methods allow for a very fast

on $\langle q \rangle$ for the 1D model with $D < n$ which has the same scaling with p, D and n as Eq. (9).

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 intend to analyze how the propagation of errors could limit the performance of the algorithm. As a hard problem with practical applications, we will be considering the Max-Cut problem. Given a graph $G = (V, E)$ on n vertices and adjacency matrix a_{ij} , the Max-Cut of G is defined as the maximization of the cost function

$$C = \frac{1}{2} \sum_{(i,j) \in E} a_{ij} (1 - Z_i Z_j), \quad (11)$$

with $Z \in \{-1, 1\}^n$. Let us denote by C_{max} the solution of the problem, and denote the average value of the cost function when random guessing by

$$C_{\text{avg}} = \text{tr}\left(C \frac{I}{2^n}\right) = \frac{1}{2} \sum_{(i,j) \in E} a_{ij}. \quad (12)$$

Since the output of the error propagation channel contains some qubits that are depolarized, on average, it will not be able to reach 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

implementation, even for large systems [43]. It is thus a reasonable assumption to make 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.

For example, we can briefly study unweighted bounded degree $\Delta = 3$ graphs, as they closely match problems of industrial interest, and 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, $C_{\text{max}} \geq 2/3|E|$ [44]. Combined with Eq. (13), this provides an upper bound for the approximation ratio,

$$\alpha \leq 1 - (\langle q \rangle / n)^2 / 4. \quad (14)$$

Furthermore, there is a classical approximation algorithm that achieves an approximation ratio of 0.9326 [45]. We therefore obtain that only when $\langle q \rangle / n \leq 0.52$ can the quantum algorithm output a better average energy than the classical one. That is, if 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. We note that this 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, it is $C_{\max} = |E|$ 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 superiority when only around 15% of the qubits are depolarized.

Using the scalings from Table I, we obtain that for shallow circuits (as defined in the table), after a depth $D = O(\sqrt{1/p})$ for 1D and $D = O(\sqrt[3]{1/p})$ for 2D 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 [34]. This is a consequence of the rapid spreading of the errors.

The impact of errors in 2D (and even all-to-all) architectures is much higher than in 1D 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 increase in the computational power of the quantum circuit with its connectivity. In order to take that into account, in the next subsection we consider a specific case, QAOA for solving a non-local problem, i.e a graph with long-range vertices (a non-planar graph). This necessarily places us in the deep circuit regime (as defined in Table I) in all cases.

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 consist usually of non-planar graphs. For example, planar Max-Cut can always be solved in polynomial time on classical computers [46]. Since non-planar 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

results. Following the scalings reported in [10, 34], we will assume that 10 layers are enough to reach the solution, and that every QAOA layer, after the routing, needs on average $\sim \sqrt{7n}$ 2-qubit gates in 2D and $3n$ 2-qubit gates in 1D. In this case, a circuit depth $D_{2D} \geq 10\sqrt{7n}$ for the 2D case is needed, and $D_{1D} \geq 30n$ in 1D. This places us in the deep circuit regime, as seen in Table I and Eqs. (9) and (10). 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 1-qubit and 2-qubit gates, thus restricting even more the error rate.

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 D 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 on not fully exploiting the whole quantum computer. Perhaps, this deserves a more careful analysis.

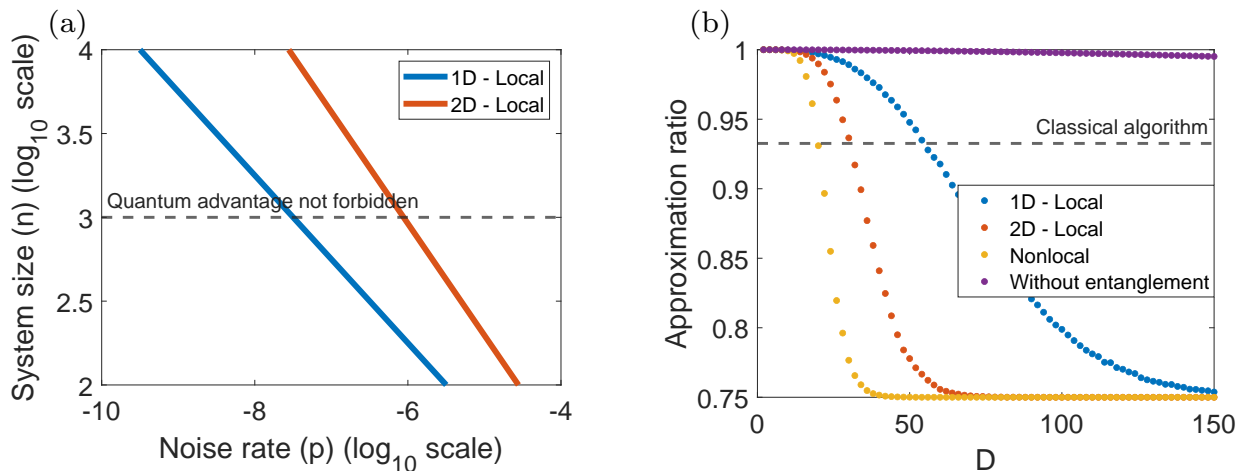


FIG. 4. In (a) we represent the relation between the system size and the error rate when the average number of depolarized qubits at the end of the computation is fixed to $\langle q \rangle/n = 0.5$ for bounded degree $\Delta = 3$ graphs. In the 1D case the depth is $D_{1D} = 30n$, and in 2D $D_{2D} = 10\sqrt{7n}$. The horizontal dashed line corresponds to system sizes of $n = 1000$, which is when a potential quantum advantage could begin to be practically useful [25]. In (b) we represent the upper bound for the approximation ratio of bounded degree $\Delta = 3$ graphs, given by Eq. (14), as a function of the circuit depth, for the different architectures: 1D, 2D, and nonlocal. We have used 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. Additionally, we have represented the upper bound on the approximation ratio that one would get when applying only local depolarizing noise to all the qubits, without any unitaries (and therefore without entanglement or propagation of errors).

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 [47]) 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 [47] 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

The 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 number 414325145 in the framework of the Austrian Science Fund (FWF): SFB F7104. RT acknowledges the Max Planck Harvard Research Center for Quantum Optics (MPHQ) postdoctoral fellowship.

[1] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando Brandao, David Buell, Brian Burkett, Yu Chen, Jimmy Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Michael Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew Harrigan, Michael Hartmann, Alan Ho, Markus Rudolf Hoffmann, Trent Huang, Travis Humble, Sergei Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, Dave Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod Ryan McClean, Matthew McEwen, Anthony Megrant, Xiao

Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin Jeffery Sung, Matt Trevithick, Amit Vainsencher, Benjamin Villalonga, Ted White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574:505–510, 2019.

[2] Yulin Wu, Wan-Su Bao, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, Cheng Guo, Chu Guo, Shaojun Guo, Lianchen Han, Linyin Hong, He-

- Liang Huang, Yong-Heng Huo, Liping Li, Na Li, Shaowei Li, Yuan Li, Futian Liang, Chun Lin, Jin Lin, Hao-ran Qian, Dan Qiao, Hao Rong, Hong Su, Lihua Sun, Liangyuan Wang, Shiyu Wang, Dachao Wu, Yu Xu, Kai Yan, Weifeng Yang, Yang Yang, Yangsen Ye, Jianghan Yin, Chong Ying, Jiale Yu, Chen Zha, Cha Zhang, Haibin Zhang, Kaili Zhang, Yiming Zhang, Han Zhao, Youwei Zhao, Liang Zhou, Qingling Zhu, Chao-Yang Lu, Cheng-Zhi Peng, Xiaobo Zhu, and Jian-Wei Pan. Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.*, 127:180501, Oct 2021.
- [3] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, Peng Hu, Xiao-Yan Yang, Wei-Jun Zhang, Hao Li, Yuxuan Li, Xiao Jiang, Lin Gan, Guangwen Yang, Lixing You, Zhen Wang, Li Li, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020.
- [4] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 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, Feb 2022.
- [6] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000.
- [7] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 90:015002, Jan 2018.
- [8] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunswoth, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, Lev Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Erik Lucero, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mroczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Hartmut Neven, Murphy Yuezhen Niu, Thomas E. O’Brien, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Doug Strain, Kevin J. Sung, Marco Szalay, Tyler Y. Takeshita, Amit Vainsencher, Theodore White, Nathan Wiebe, Z. Jamie Yao, Ping Yeh, and Adam Zalcman. Hartree-fock on a superconducting qubit quantum computer. *Science*, 369(6507):1084–1089, 2020.
- [9] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.
- [10] Matthew P Harrigan, Kevin J Sung, Matthew Neeley, Kevin J Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C Bardin, Rami Barends, Sergio Boixo, et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, 2021.
- [11] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, feb 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, Jun 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. *Nature communications*, 5(1):1–7, 2014.
- [14] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Phys. Rev. A*, 92:042303, Oct 2015.
- [15] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse ising model. *Phys. Rev. E*, 58:5355–5363, Nov 1998.
- [16] Suguru Endo, Iori Kurata, and Yuya O. Nakagawa. Calculation of the green’s function on near-term quantum computers. *Phys. Rev. Research*, 2:033281, Aug 2020.
- [17] A.B. Finnila, M.A. Gomez, C. Sebenik, C. Stenson, and J.D. Doll. Quantum annealing: A new method for minimizing multidimensional functions. *Chemical Physics Letters*, 219(5):343–348, 1994.
- [18] Hongbin Liu, Guang Hao Low, Damian S Steiger, Thomas Häner, Markus Reiher, and Matthias Troyer. Prospects of quantum computing for molecular sciences. *Materials Theory*, 6(1):1–17, 2022.
- [19] Andrew A Houck, Hakan E Türeci, and Jens Koch. On-chip quantum simulation with superconducting circuits. *Nature Physics*, 8(4):292–299, 2012.
- [20] 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, Sep 2021.
- [21] Bernhard H Korte, Jens Vygen, B Korte, and J Vygen. *Combinatorial optimization*, volume 1. Springer, 2011.
- [22] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.
- [23] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm. *arXiv preprint arXiv:1602.07674*, 2016.
- [24] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou. The quantum approximate optimization algorithm and the sherrington-kirkpatrick model at infinite size. *arXiv preprint arXiv:1910.08187*, 2019.
- [25] Gian Giacomo Guerreschi and Anne Y Matsuura. Qaoa for max-cut requires hundreds of qubits for quantum speed-up. *Scientific reports*, 9(1):1–7, 2019.
- [26] 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, Jun 2020.

- [27] 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. *Proceedings of the National Academy of Sciences*, 117(41):25396–25401, 2020.
- [28] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications*, 12(1):1–12, 2021.
- [29] Tobias Haug, Kishor Bharti, and M.S. Kim. Capacity and quantum geometry of parametrized quantum circuits. *PRX Quantum*, 2:040309, Oct 2021.
- [30] Kouhei Nakaji and Naoki Yamamoto. Expressibility of the alternating layered ansatz for quantum computation. *Quantum*, 5:434, Apr 2021.
- [31] V. Akshay, H. Philathong, M. E. S. Morales, and J. D. Biamonte. Reachability deficits in quantum approximate optimization. *Phys. Rev. Lett.*, 124:090504, Mar 2020.
- [32] Jeffrey Marshall, Filip Wudarski, Stuart Hadfield, and Tad Hogg. Characterizing local noise in QAOA circuits. *IOP SciNotes*, 1(2):025208, aug 2020.
- [33] Cheng Xue, Zhao-Yun Chen, Yu-Chun Wu, and Guo-Ping Guo. Effects of quantum noise on quantum approximate optimization algorithm. *Chinese Physics Letters*, 38(3):030302, 2021.
- [34] Daniel Stilck França and Raul Garcia-Patron. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, 2021.
- [35] Dorit Aharonov, Michael Ben-Or, Russell Impagliazzo, and Noam Nisan. Limitations of noisy reversible computation. *arXiv preprint quant-ph/9611028*, 1996.
- [36] Michael Ben-Or, Daniel Gottesman, and Avinatan Hassidim. Quantum refrigerator. *arXiv preprint arXiv:1301.1995*, 2013.
- [37] Adam Bouldan, Bill Fefferman, Zeph Landau, and Yunchao Liu. Noise and the frontier of quantum supremacy. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1308–1317. IEEE, 2022.
- [38] 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. *arXiv preprint arXiv:2112.00716*, 2021.
- [39] Alexander M Dalzell, Nicholas Hunter-Jones, and Fernando GSL Brandão. Random quantum circuits transform local noise into global white noise. *arXiv preprint arXiv:2111.14907*, 2021.
- [40] 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. *Nature Physics*, 14(6):595–600, 2018.
- [41] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347–S352, sep 2005.
- [42] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, nov 1995.
- [43] Alp Yurtsever, Joel A. Tropp, Olivier Fercoq, Madeleine Udell, and Volkan Cevher. Scalable semidefinite programming. *SIAM Journal on Mathematics of Data Science*, 3(1):171–200, 2021.
- [44] Paul Erdős. On some extremal problems in graph theory. *Israel Journal of Mathematics*, 3(2):113–116, 1965.
- [45] Eran Halperin, Dror Livnat, and Uri Zwick. Max cut in cubic graphs. *Journal of Algorithms*, 53(2):169–185, 2004.
- [46] F. Hadlock. Finding a maximum cut of a planar graph in polynomial time. *SIAM Journal on Computing*, 4(3):221–225, 1975.
- [47] Benjamin F Schiffer, Jordi Tura, and J Ignacio Cirac. Adiabatic spectroscopy and a variational quantum adiabatic algorithm. *arXiv preprint arXiv:2103.01226*, 2021.
- [48] Gregory F Lawler. *Introduction to stochastic processes*. Chapman and Hall/CRC, 2018.
- [49] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- [50] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357 – 367, 1967.

Appendix A: Proof of energy bound

Here we derive the energy bound given in Eq. (13). As explained in the main text, we are considering the Max-Cut problem. Given a graph $G = (V, E)$ on n vertices and adjacency matrix a_{ij} the Max-Cut of G is defined as the maximization of the cost function.

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

with $Z \in \{-1, 1\}^n$. Let us denote by C_{\max} the solution of the problem, and denote the average value of the cost function when random guessing by

$$C_{\text{avg}} = \text{tr} \left(C \frac{I}{2^n} \right) = \frac{1}{2} \sum_{(i,j) \in E} a_{ij} \quad (\text{A2})$$

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. As discussed in Eq. (5), our model outputs a state where some qubits are in their correct state and others are depolarized. We denote by $P(Q_i)$ the probability that qubit i is in the correct value. While we do not have access to the complete probability distribution, we know that $P(Q_i) = r$ is the same for all qubits, and that $P(Q_i \cap Q_j) \geq r^2$ for all pairs (i, j) . On the graph, we assume that $a_{ij} \geq 0 \forall (i, j)$. 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 are random, after averaging the contribution is always $a_{ij}/2$, since it will be cut half of the times. Then, the total cost will be:

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) = \sum_{(i,j) \in M} a_{ij} P(Q_i \cap Q_j) + \frac{1}{2} \sum_{(i,j) \in E} a_{ij} (1 - P(Q_i \cap Q_j)) \quad (\text{A3})$$

Rearranging the terms this yields:

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) = \frac{1}{2} \sum_{(i,j) \in M} a_{ij} (1 + P(Q_i \cap Q_j)) + \frac{1}{2} \sum_{(i,j) \in E \setminus M} a_{ij} (1 - P(Q_i \cap Q_j)) \quad (\text{A4})$$

We can now use the fact that $P(Q_i \cap Q_j) \leq P(Q_i) = r$, and $P(Q_i \cap Q_j) \geq r^2$, to obtain:

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) \leq \frac{1}{2} (1+r) \sum_{(i,j) \in M} a_{ij} + \frac{1}{2} (1-r^2) \sum_{(i,j) \in E \setminus M} a_{ij} = \frac{1}{2} (1+r) C_{\text{max}} + (1-r^2) \left(C_{\text{avg}} - \frac{C_{\text{max}}}{2} \right) \quad (\text{A5})$$

Then, since $r = 1 - \langle q \rangle / n$, we obtain:

$$\text{tr}(C\Phi_{\text{avg}}(\rho)) \leq \frac{1}{2} \left(1 - \frac{\langle q \rangle}{n} \right) \left(2 - \frac{\langle q \rangle}{n} \right) C_{\text{max}} + \left(1 - \left(1 - \frac{\langle q \rangle}{n} \right)^2 \right) C_{\text{avg}} \quad (\text{A6})$$

Appendix B: Heuristic formula

We will explain here how we obtained the heuristic formula for the 1D case, in Eq. (9) from the main text. 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 the main text. 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 $M_t = M_t^{\text{noise}} M_t^U$. As explained in the definition, D_M is the number of times that we apply the Markov chain, and it takes the value $D_M = D/2$, where D is the circuit depth of the associated quantum channel. The state space of the Markov chain comprises of strings of zeroes and ones, $\{0, 1\}^n$. Mapping this to the outcome of the quantum channel Φ_{avg} , the zeroes in the string determine which qubits are in the correct state, while the ones determine which qubits are depolarized. We will denote the average number of ones at a given time step t as $\langle q(t) \rangle$.

In every step there are therefore two matrices: M_{noise} , which applies the noise with probability p_M , and M_U , which propagates it. As a good approximation, here we will take $p_M = 2p$, where p is the error rate of the quantum channel. Since we are in 1D, we are applying layers of unitaries to all the even (or odd) pairs, and M_U will just consist in applying the matrix P 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 will just apply M_U from then on. Since M_U can only propagate this error to neighbouring 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. Since we know the expression of M_U , it is easy to see that it is a one-dimensional lazy random walk on the line.

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

$$P(x, y) = \begin{cases} \left(\frac{4}{5}\right)^2 & \text{if } y = x - 2 \\ \left(\frac{1}{5}\right)^2 & \text{if } y = x + 2 \\ 2 \left(\frac{4}{5}\right) \left(\frac{1}{5}\right) & \text{if } y = x \\ 0 & \text{else,} \end{cases} \quad (\text{B1})$$

for $x \in (2, n-2)$. In the edges, $P(0,0) = 1$, $P(n,n) = 1$, $P(0,1) = 1/5$, $P(2,1) = 4/5$, $P(n,n-1) = 4/5$, $P(n-2,n-1) = 1/5$.

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

Lemma 2 (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 X_t the state after t applications of the Markov chain:*

$$P(X_t = 0 | X_0 = 1) \leq \frac{1}{4}, \forall t. \quad (\text{B2})$$

Proof: This can be shown using the standard techniques for random walks with absorbing barriers [48]. The random walk starts in the 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

$$\left(\begin{array}{c|c} I & S \\ \hline 0 & Q \end{array} \right).$$

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

$$\lim_{t \rightarrow \infty} P(X_t = 0 | X_0 = 2) = (1/5)^2 \left[\frac{(16/25)^n - (1/25)^n}{(16/25)^{n+1} - (1/25)^{n+1}} \right] \leq \frac{1}{16}. \quad (\text{B3})$$

Then:

$$\lim_{t \rightarrow \infty} P(X_t = 0 | X_0 = 1) \leq \frac{1}{5} + \frac{4}{5} \frac{1}{16} = \frac{1}{4}. \quad (\text{B4})$$

□

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 1/4 of the times, 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

$$A = \int_0^{D_M} \frac{3}{4} \min(6/5t, n) = H\left(D_M - \frac{5}{6}n\right) \left(\frac{3}{4}nD_M - \frac{5}{16}n^2\right) + H\left(-D_M + \frac{5}{6}n\right) \left(\frac{9}{20}D_M^2\right), \quad (\text{B5})$$

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 3 *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:*

$$x_k = \begin{cases} 1 & \text{if } |j-k| < 3/4 \min(6/5(t-t^*), n) \\ 0 & \text{if } |j-k| > 3/4 \min(6/5(t-t^*), n). \end{cases} \quad (\text{B6})$$

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. (B5). 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

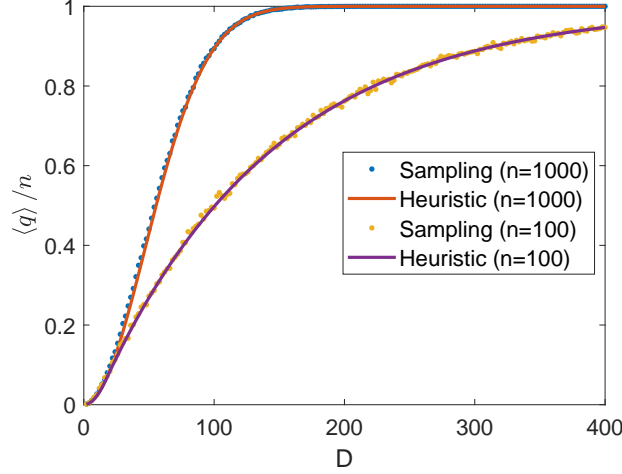


FIG. 5. The heuristic formula in Eq. (9) is plotted along with the results from sampling from the Markov chain for the 1D architecture. This is done for an error rate $p = 10^{-3}$ and with two different system sizes, $n = 100$ and $n = 1000$. The heuristic formula shows good agreement with the result from sampling from the Markov chain.

that this model provides a good approximation of the average number of ones in the Markov chain. Setting $p_M = 2p$ and $D_M = D/2$ gives the final expression:

$$\langle q \rangle_{1D} / n \simeq \begin{cases} 1 - (1 - 2p)^{\frac{9}{80}D^2} & \text{if } D \leq \frac{5}{3}n \\ 1 - (1 - 2p)^{\frac{3}{8}nD - \frac{5}{16}n^2} & \text{if } D > \frac{5}{3}n \end{cases} \quad (\text{B7})$$

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 C: Rigorous 1D bound

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

We are working with the Markov chain as defined in the main text (Eq. (7)). In this picture, an error in qubit k corresponds to flipping bit k from 0 to 1. We define the set of 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 $P(\overline{Q}_j)$ the probability that bit j is in 1 at the end of the computation. This is given by

$$P(\overline{Q}_j) = \sum_{s \subset \mathcal{S}} P(\overline{Q}_j | s) P(s). \quad (\text{C1})$$

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

Let us now assume that there exists a subset of errors $A \subset \mathcal{S}$ such that, $P(\overline{Q}_j | (a, b) \in A) \geq c$, 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

$$\sum_s P(\overline{Q}_j | s) P(s) \geq \sum_s P(\overline{Q}_j | (a, b) \in s) P(s) \geq \sum_s P(s \cap A \neq \emptyset) c = c \left(\sum_s 1 - P(s \cap A = \emptyset) \right) = c \left[1 - (1 - p_M)^{|A|} \right]. \quad (\text{C2})$$

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 $P(\overline{Q}_j | (k, t^*))$. As explained in appendix B, 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 following random variables*

$$X_A(t) = X_A^0 + \sum_{i=1}^{t^*} X_A^{(i)}, \quad (\text{C3})$$

$$X_B(t) = X_B^0 + \sum_{i=1}^{t^*} X_B^{(i)}, \quad (\text{C4})$$

where $X_A^0 = k$, and

$$X_B^0 = \begin{cases} k + 1 & \text{with probability } 4/5 \\ k - 1 & \text{with probability } 1/5, \end{cases} \quad (\text{C5})$$

$$X_A^{(i)} = \begin{cases} 0 & \text{if } X_A(t-1) > X_B(t-1) \\ 1 & \text{with probability } 1/5 \text{ if } X_A(t-1) < X_B(t-1) \\ -1 & \text{with probability } 4/5 \text{ if } X_A(t-1) < X_B(t-1), \end{cases} \quad (\text{C6})$$

$$X_B^{(i)} = \begin{cases} 0 & \text{if } X_A(t-1) > X_B(t-1) \\ -1 & \text{with probability } 1/5 \text{ if } X_A(t-1) < X_B(t-1) \\ 1 & \text{with probability } 4/5 \text{ if } X_A(t-1) < X_B(t-1). \end{cases} \quad (\text{C7})$$

We also define two independent random walkers as

$$X_A^{\text{ind}}(t) = X_A^0 + \sum_{i=1}^{t^*} X_A^{\text{ind},(i)}, \quad (\text{C8})$$

$$X_B^{\text{ind}}(t) = X_B^0 + \sum_{i=1}^{t^*} X_B^{\text{ind},(i)}, \quad (\text{C9})$$

where

$$X_A^{\text{ind},(i)} = \begin{cases} 1 & \text{with probability } 1/5 \\ -1 & \text{with probability } 4/5, \end{cases} \quad (\text{C10})$$

$$X_B^{\text{ind},(i)} = \begin{cases} -1 & \text{with probability } 1/5 \\ 1 & \text{with probability } 4/5. \end{cases} \quad (\text{C11})$$

With the definitions above, it holds that

$$P(\overline{Q_j}|(k, t^*)) = P(X_B \geq j \cap X_A \leq j). \quad (\text{C12})$$

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 on the rest.

Lemma 4 *Let us consider the random variables X_A , X_A^{ind} , X_B and X_B^{ind} as defined in definition 1. Then,*

$$P(\overline{Q_j}|(k, t^*)) = P(X_B \geq j \cap X_A \leq j) \geq P(X_B^{\text{ind}} \geq j) P(X_A^{\text{ind}} \leq j) - \frac{1}{4}. \quad (\text{C13})$$

Proof: We want to compute $P(X_B \geq j \cap X_A \leq j)$. 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 $D_M < n/2$, from Lemma 2 we know that the probability that they cross is upper bounded, $P(C) \leq 1/4$. Then, we can write:

$$P(X_A \geq j \cap X_B \leq j) = P(X_A \geq j \cap X_B \leq j | NC) P(NC). \quad (\text{C14})$$

We can now consider the two random walks X_A^{ind} and X_B^{ind} . Their behavior is the same as X_A and X_B provided there is no crossing. That is:

$$P(X_A \geq j \cap X_B \leq j | NC) = P(X_A^{\text{ind}} \geq j \cap X_B^{\text{ind}} \leq j | NC). \quad (\text{C15})$$

Using now the law of total probability we get:

$$\begin{aligned} P(X_A^{\text{ind}} \geq j \cap X_B^{\text{ind}} \leq j | NC) P(NC) = \\ P(X_A^{\text{ind}} \geq j) P(X_B^{\text{ind}} \leq j) - P(X_A^{\text{ind}} \geq j \cap X_B^{\text{ind}} \leq j | C) P(C) \geq \\ P(X_A^{\text{ind}} \geq j) P(X_B^{\text{ind}} \leq j) - \frac{1}{4}. \end{aligned} \quad (\text{C16})$$

The result then follows immediately from Eq. (C12). \square

We know now how to bound the probability $P(\overline{Q_j}|(k, t^*))$ using two biased random walks on the line. This allows us to bound the region:

Lemma 5 *Let us consider an error (k, t^*) such that*

$$|j - k| \leq \frac{6}{5}(D_M - t^*) - 2\sqrt{2(D_M - t^*) \ln \frac{1}{3/4 - c}}. \quad (\text{C17})$$

Then, $P(\overline{Q_j}|(k, t^*)) \geq c$.

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 [49]:

$$P(X_A^{\text{ind}} - E[X_A^{\text{ind}}] \geq \alpha) \leq \exp\left(-\frac{\alpha^2}{2(D_M - t^*)}\right). \quad (\text{C18})$$

The expectation values can easily be computed as $E[X_A^{\text{ind}}] = k - 3/5(D_M - t^*)$ and $E[X_B^{\text{ind}}] = k + 3/5(1 + D_M - t^*)$

Let us now assume, without loss of generality, that $j < k$. We want to compute the quantity $P(X_A^{\text{ind}} \geq j) P(X_B^{\text{ind}} \leq j)$. We have

$$P(X_A^{\text{ind}} \leq j) = 1 - P(X_A^{\text{ind}} \geq j) = 1 - P(X_A^{\text{ind}} - E[X_A^{\text{ind}}] \geq j - E[X_A^{\text{ind}}]) \geq 1 - \exp\left(-\frac{(j - E[X_A^{\text{ind}}])^2}{2(D_M - t^*)}\right), \quad (\text{C19})$$

$$P(X_B^{\text{ind}} \geq j) \geq 1 - \exp\left(-\frac{(j - E[X_B^{\text{ind}}])^2}{2(D_M - t^*)}\right). \quad (\text{C20})$$

We note that, since $j < k$, the dominant term is $P(X_A^{\text{ind}} \geq j)$, and therefore we neglect the term with X_B , which is exponentially smaller. Setting $P(Q_j | (k, t^*)) \geq P(X_A^{\text{ind}} \geq j) - 1/4 \geq c$ we directly obtain

$$\left|j - k + \frac{3}{5}(D_M - t^*)\right| \leq \sqrt{2(D_M - t^*) \ln \frac{1}{3/4 - c}}. \quad (\text{C21})$$

Doing the same for the cases where $j < k$, which corresponds to the region where the term with X_B will dominate, completes the computation. \square

We have therefore found a region such that $P(\overline{Q}_j | (k, t^*)) \geq c$. Integration over t^* yields:

$$|A| = \frac{3}{5}D_M^2 - \frac{4}{3}\sqrt{2 \ln \frac{1}{3/4 - c}} D_M^{3/2}. \quad (\text{C22})$$

Then, using Eq. (C2) we can bound

$$\langle q \rangle / n \geq c \left(1 - (1 - p_M)^{\frac{3}{5}D_M^2 - \frac{4}{3}\sqrt{2 \ln \frac{1}{3/4 - c}} D_M^{3/2}}\right). \quad (\text{C23})$$

As conjectured, the number of depolarized qubits scales exponentially with $O(D^2)$.

Appendix D: Proof of 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 the 1D case 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].

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, $X_i = \text{tr}(\rho C_i)$, corresponding to each of the edges of the graph. 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 $|Z_i - Z_{i-1}|$, 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 $|Z_i - Z_{i-1}| \leq \Delta D$. Applying Azuma-Hoeffding's inequality yields:

$$\Pr[|C - \langle C \rangle| \geq \lambda] \leq 2e^{-\frac{\lambda^2}{2|E|\Delta^2 D^2}}. \quad (\text{D1})$$

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

$$\Pr[|C - \langle C \rangle| \geq \alpha|E|] \leq 2e^{-\frac{\alpha^2|E|}{2\Delta^2 D^2}}. \quad (\text{D2})$$

As a consequence, for constant depth, the value of C is concentrated around its mean, $\langle C \rangle$.