

Noise-induced barren plateaus in variational quantum algorithms

Samson Wang ^{1,2}, Enrico Fontana^{1,3,4}, M. Cerezo ^{1,5}, Kunal Sharma^{1,6,7}, Akira Sone ^{1,5,8}, Lukasz Cincio¹ & Patrick J. Coles ¹

Variational Quantum Algorithms (VQAs) may be a path to quantum advantage on Noisy Intermediate-Scale Quantum (NISQ) computers. A natural question is whether noise on NISQ devices places fundamental limitations on VQA performance. We rigorously prove a serious limitation for noisy VQAs, in that the noise causes the training landscape to have a barren plateau (i.e., vanishing gradient). Specifically, for the local Pauli noise considered, we prove that the gradient vanishes exponentially in the number of qubits n if the depth of the ansatz grows linearly with n . These noise-induced barren plateaus (NIBPs) are conceptually different from noise-free barren plateaus, which are linked to random parameter initialization. Our result is formulated for a generic ansatz that includes as special cases the Quantum Alternating Operator Ansatz and the Unitary Coupled Cluster Ansatz, among others. For the former, our numerical heuristics demonstrate the NIBP phenomenon for a realistic hardware noise model.

¹Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ²Imperial College London, London, UK. ³University of Strathclyde, Glasgow, UK. ⁴National Physical Laboratory, Teddington, UK. ⁵Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA.

⁶Hearne Institute for Theoretical Physics and Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA, USA. ⁷Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, Maryland, MD, USA. ⁸Aliro Technologies, Inc, Boston, MA 02135, USA.

email: samsonwang@outlook.com; cerezo@lanl.gov; pcoles@lanl.gov

One of the great unanswered technological questions is whether Noisy Intermediate Scale Quantum (NISQ) computers will yield a quantum advantage for tasks of practical interest¹. At the heart of this discussion are Variational Quantum Algorithms (VQAs), which are believed to be the best hope for near-term quantum advantage^{2–4}. Such algorithms leverage classical optimizers to train the parameters in a quantum circuit, while employing a quantum device to efficiently estimate an application-specific cost function or its gradient. By keeping the quantum circuit depth relatively short, VQAs mitigate hardware noise and may enable near-term applications including electronic structure^{5–8}, dynamics^{9–12}, optimization^{13–16}, linear systems^{17,18}, metrology^{19,20}, factoring²¹, compiling^{22–24}, and others^{25–30}.

The main open question for VQAs is their scalability to large problem sizes. While performing numerical heuristics for small or intermediate problem sizes is the norm for VQAs, deriving analytical scaling results is rare for this field. Noteworthy exceptions are some recent studies of the scaling of the gradient in VQAs with the number of qubits n ^{31–39}. For example, it was proven that the gradient vanishes exponentially in n for randomly initialized, deep Hardware Efficient ansatzes^{31,32} and dissipative quantum neural networks³³, and also for shallow depth with global cost functions³⁴. This vanishing gradient phenomenon is now referred to as barren plateaus in the training landscape. Barren plateaus imply that in order to resolve gradients to a fixed precision, on average, an exponential number of shots need to be invested. This places an exponential resource burden on the training process of VQAs. Further, such effects are not avoided by adopting optimizers that use information about higher order derivatives³⁸ or gradient-free methods³⁹. Fortunately, investigations into barren plateaus have spawned several promising strategies to avoid them, including local cost functions^{34,40}, parameter correlation³⁷, pre-training⁴¹, and layer-by-layer training^{42,43}. Such strategies give hope that perhaps VQAs may avoid the exponential scaling that otherwise would result from the exponential precision requirements of navigating through a barren plateau.

However, these works do not consider quantum hardware noise, and very little is known about the scalability of VQAs in the presence of noise. One of the main selling points of VQAs is noise mitigation, and indeed VQAs have shown evidence of optimal parameter resilience to noise in the sense that the global minimum of the landscape may be unaffected by noise^{6,23}. While some analysis has been done^{44–46}, an important open question, which has not yet been addressed, is how noise affects the asymptotic scaling of VQAs. More specifically, one can ask how noise affects the training process. If the effect of noise on trainability is not severe, and the optimal parameters can be found, then VQAs may be useful even in the presence of high decoherence in one of two ways. First, the end goal of certain algorithms such as the Quantum Approximate Optimization Algorithm (QAOA)⁴⁷ is to extract an optimized set of parameters, rather than the optimal cost value. Second, error mitigation can be used in conjunction with VQAs that display optimal parameter resilience. Intuitively, incoherent noise is expected to reduce the magnitude of the gradient and hence hinder trainability, and preliminary numerical evidence of this has been seen^{48,49}, although the scaling of this effect has not been studied.

In this work, we analytically study the scaling of gradient for VQAs as a function of n , the circuit depth L , and a noise parameter $q < 1$. We consider a general class of local noise models that includes depolarizing noise and certain kinds of Pauli noise. Furthermore, we investigate a general, abstract ansatz that allows us to encompass many of the important ansatzes in the literature,

hence allowing us to make a general statement about VQAs. This includes the Quantum Alternating Operator Ansatz (QAOA) which is used for solving combinatorial optimization problems^{13–16} and the Unitary Coupled Cluster (UCC) Ansatz which is used in the Variational Quantum Eigensolver (VQE) to solve chemistry problems^{50–52}. This is also applicable for the Hardware Efficient Ansatz and the Hamiltonian Variational Ansatz (HVA) which are employed for various applications^{53–57}. Our results also generalize to settings that allow for multiple input states or training data, as in machine learning applications, often called quantum neural networks^{58–62}.

Our main result (Theorem 1) is an upper bound on the magnitude of the gradient that decays exponentially with L , namely as $2^{-\kappa}$ with $\kappa = -L \log_2(q)$. Hence, we find that the gradient vanishes exponentially in the circuit depth. Moreover, it is typical to consider L scaling as poly(n) (e.g., in the UCC Ansatz⁵²), for which our main result implies an exponential decay of the gradient in n . We refer to this as a Noise-Induced Barren Plateau (NIBP). We remark that NIBPs can be viewed as concomitant to the cost landscape concentrating around the value of the cost for the maximally mixed state, and we make this precise in Lemma 1. See Fig. 1 for a schematic diagram of the NIBP phenomenon.

To be clear, any variational algorithm with a NIBP will have exponential scaling. In this sense, NIBPs destroy quantum speedup, as the standard goal of quantum algorithms is to avoid the typical exponential scaling of classical algorithms. NIBPs are conceptually distinct from the noise-free barren plateaus of refs. ^{31–36}. Indeed, strategies to avoid noise-free barren plateaus^{34,37,40–43} do not appear to solve the NIBPs issue.

The obvious strategy to address NIBPs is to reduce circuit complexity, or more precisely, to reduce the circuit depth. Hence, our work provides quantitative guidance for how small L needs to be to potentially avoid NIBPs.

In what follows, we present our general framework followed by our main result. We also present two extensions of our main result, one involving correlated ansatz parameters and one allowing for measurement noise. The latter indicates that global cost functions exacerbate the NIBP issue. In addition, we provide numerical heuristics that illustrate our main result for MaxCut optimization with the QAOA, and an implementation of the HVA on superconducting hardware, both showing that NIBPs significantly impact this application.

Results

General framework. In this work we analyze a general class of parameterized ansatzes $U(\theta)$ that can be expressed as a product of L unitaries sequentially applied by layers

$$U(\theta) = U_L(\theta_L) \cdots U_2(\theta_2) \cdot U_1(\theta_1). \quad (1)$$

Here $\theta = \{\theta_i\}_{i=1}^L$ is a set of vectors of continuous parameters that are optimized to minimize a cost function C that can be expressed as the expectation value of an operator O :

$$C = \text{Tr}[OU(\theta)\rho U^\dagger(\theta)]. \quad (2)$$

As shown in Fig. 2, ρ is an n -qubit input state. Without loss of generality we assume that each $U_i(\theta_i)$ is given by

$$U_i(\theta_i) = \prod_m e^{-i\theta_{im}H_{im}} W_{lm}, \quad (3)$$

where H_{im} are Hermitian operators, $\theta_i = \{\theta_{im}\}$ are continuous parameters, and W_{lm} denote unparametrized gates. We expand H_{im} and O in the Pauli basis as

$$H_{lm} = \eta_{lm} \cdot \sigma_n = \sum_i \eta_{lm}^i \sigma_n^i, \quad O = \omega \cdot \sigma_n = \sum_i \omega^i \sigma_n^i, \quad (4)$$

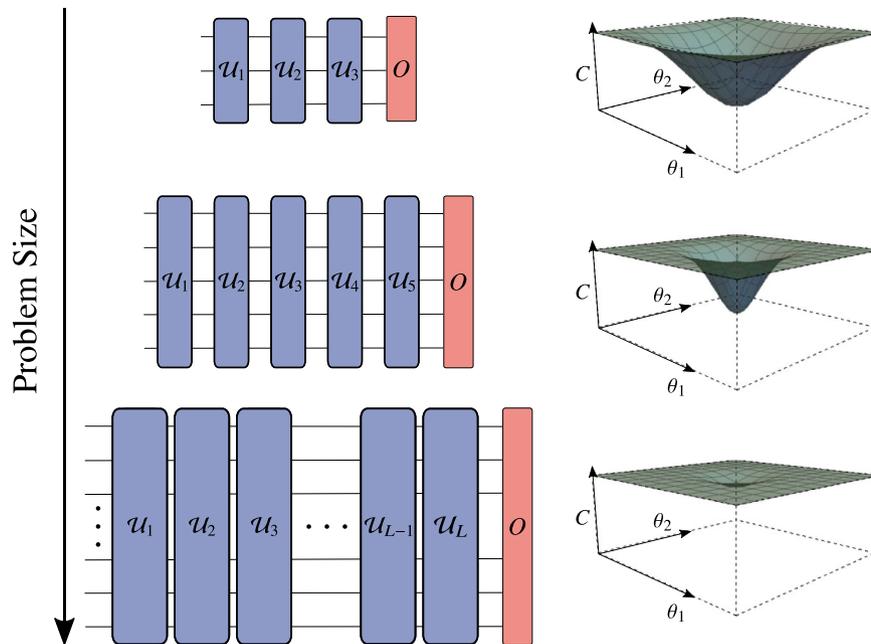


Fig. 1 Schematic diagram of the Noise-Induced Barren Plateau (NIBP) phenomenon. For various applications such as chemistry and optimization, increasing the problem size often requires one to increase the depth L of the variational ansatz. We show that, in the presence of local noise, the gradient vanishes exponentially in L and hence exponentially in the number of qubits n when L scales linearly in n . This can be seen in the plots on the right, which show the cost function landscapes for a simple variational problem with local noise.

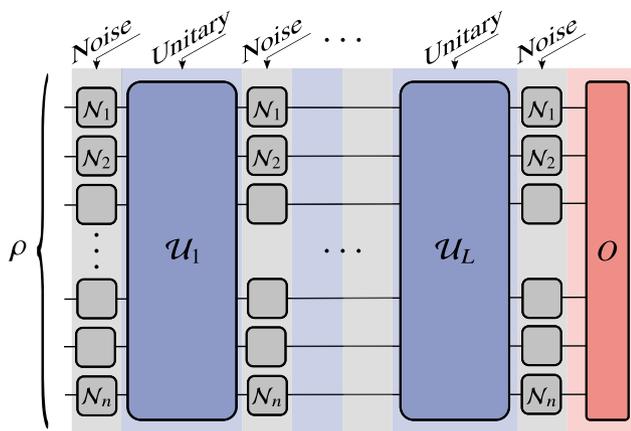


Fig. 2 Setting for our analysis. An n -qubit input state ρ is sent through a variational ansatz $U(\theta)$ composed of L unitary layers $U_i(\theta_i)$ sequentially acting according to Eq. (1). Here, U_i denotes the quantum channel that implements the unitary $U_i(\theta_i)$. The parameters in the ansatz $\theta = \{\theta_i\}_{i=1}^L$ are trained to minimize a cost function that is expressed as the expectation value of an operator O as in Eq. (2). We consider a noise model where local Pauli noise channels \mathcal{N}_j act on each qubit j before and after each unitary.

where $\sigma_n^j \in \{\mathbb{1}, X, Y, Z\}^{\otimes n}$ are Pauli strings, and η_{lm} and ω are real-valued vectors that specify the terms present in the expansion. Defining $N_{lm} = |\eta_{lm}|$ and $N_O = |\omega|$ as the number of non-zero elements, i.e., the number of terms in the summations in Eq. (4), we say that H_{lm} and O admit an efficient Pauli decomposition if $N_{lm}, N_O \in \mathcal{O}(\text{poly}(n))$, respectively.

We now briefly discuss how the QAOA, UCC, and Hardware Efficient ansatzes fit into this general framework. We refer the reader to the Methods for additional details. In QAOA one sequentially alternates the action of two unitaries as

$$U(\gamma, \beta) = e^{-i\beta_p H_M} e^{-i\gamma_p H_P} \dots e^{-i\beta_1 H_M} e^{-i\gamma_1 H_P}, \quad (5)$$

where H_P and H_M are the so-called problem and mixer Hamiltonian, respectively. We define $N_P(N_M)$ as the number of terms in the Pauli decomposition of $H_P(H_M)$. On the other hand, Hardware Efficient ansatzes naturally fit into Eqs. (1)–(3) as they are usually composed of fixed gates (e.g. controlled NOTs), and parametrized gates (e.g., single qubit rotations). Finally, as detailed in the Methods, the UCC ansatz can be expressed as

$$U(\theta) = \prod_{lm} U_{lm}(\theta_{lm}) = \prod_{lm} e^{i\theta_{lm} \sum_i \mu_{lm}^i \sigma_n^i}, \quad (6)$$

where $\mu_{lm}^i \in \{0, \pm 1\}$, and where θ_{lm} are the coupled cluster amplitudes. Moreover, we denote $\tilde{N}_{lm} = |\mu_{lm}|$ as the number of non-zero elements in $\sum_i \mu_{lm}^i \sigma_n^i$.

As shown in Fig. 2, we consider a noise model where local Pauli noise channels \mathcal{N}_j act on each qubit j before and after each unitary $U_i(\theta_i)$. The action of \mathcal{N}_j on a local Pauli operator $\sigma \in \{X, Y, Z\}$ can be expressed as

$$\mathcal{N}_j(\sigma) = q_\sigma \sigma, \quad (7)$$

where $-1 < q_X, q_Y, q_Z < 1$. Here, we characterize the noise strength with a single parameter $q = \sqrt{\max\{|q_X|, |q_Y|, |q_Z|\}}$. Let \mathcal{U}_i denote the channel that implements the unitary $U_i(\theta_i)$ and let $\mathcal{N} = \mathcal{N}_1 \otimes \dots \otimes \mathcal{N}_n$ denote the n -qubit noise channel. Then, the noisy cost function is given by

$$\tilde{C} = \text{Tr}[O (\mathcal{N} \circ \mathcal{U}_L \circ \dots \circ \mathcal{N} \circ \mathcal{U}_1 \circ \mathcal{N})(\rho)]. \quad (8)$$

General analytical results. There are some VQAs, such as the VQE⁵ for chemistry and other physical systems, where it is important to accurately characterize the value of the cost function itself. We provide an important result below in Lemma 1 that quantitatively bounds the cost function itself, and we envision that this bound will be especially useful in the context of VQE. On the other hand, there are other VQAs, such as those for optimization^{13–16}, compiling^{22–24}, and linear systems^{17,18}, where the key goal is to learn the optimal parameters and the precise

value of the cost function is either not important or can be computed classically after learning the parameters. In this case, one is primarily concerned with trainability, and hence the gradient is a key quantity of interest. These applications motivate our main result in Theorem 1, which bounds the magnitude of the gradient. We remark that trainability is of course also important for VQE, and hence Theorem 1 is also of interest for this application.

With this motivation in mind, we now present our main results. We first present our bound on the cost function, since one can view this as a phenomenon that naturally accompanies our main theorem. Namely, in the following lemma, we show that the noisy cost function concentrates around the corresponding value for the maximally mixed state.

Lemma 1. (Concentration of the cost function). Consider an L -layered ansatz of the form in Eq. (1). Suppose that local Pauli noise of the form of Eq. (7) with noise strength q acts before and after each layer as in Fig. 2. Then, for a cost function \tilde{C} of the form in Eq. (8), the following bound holds

$$\left| \tilde{C} - \frac{1}{2^n} \text{Tr}[O] \right| \leq G(n) \left\| \rho - \frac{\mathbb{1}}{2^n} \right\|_1, \quad (9)$$

where

$$G(n) = N_O \|\omega\|_\infty q^{2L+2}. \quad (10)$$

Here $\|\cdot\|_\infty$ is the infinity norm, $\|\cdot\|_1$ is the trace norm, ω is defined in Eq. (4), and $N_O = |\omega|$ is the number of non-zero elements in the Pauli decomposition of O .

This lemma implies the cost landscape exponentially concentrates on the value $\text{Tr}[O]/2^n$ for large n , whenever the number of layers L scales linearly with the number of qubits. While this lemma has important applications on its own, particularly for VQE, it also provides intuition for the NIBP phenomenon, which we now state.

Let $\partial_{\theta_{lm}} \tilde{C} = \partial \tilde{C} / \partial \theta_{lm}$ denote the partial derivative of the noisy cost function with respect to the m -th parameter that appears in the l -th layer of the ansatz, as in Eq. (3). For our main technical result, we upper bound $|\partial_{\theta_{lm}} \tilde{C}|$ as a function of L and n .

Theorem 1. (Upper bound on the partial derivative). Consider an L -layered ansatz as defined in Eq. (1). Let θ_{lm} denote the trainable parameter corresponding to the Hamiltonian H_{lm} in the unitary $U_l(\theta_l)$ appearing in the ansatz. Suppose that local Pauli noise of the form in Eq. (7) with noise parameter q acts before and after each layer as in Fig. 2. Then the following bound holds for the partial derivative of the noisy cost function

$$|\partial_{\theta_{lm}} \tilde{C}| \leq F(n), \quad (11)$$

where

$$F(n) = \sqrt{8 \ln 2} N_O \|H_{lm}\|_\infty \|\omega\|_\infty n^{1/2} q^{L+1}, \quad (12)$$

and ω is defined in Eq. (4), with number of non-zero elements N_O .

Let us now consider the asymptotic scaling of the function $F(n)$ in Eq. (12). Under standard assumptions such as that O in Eq. (4) admits an efficient Pauli decomposition and that H_{lm} has bounded eigenvalues, we now state that $F(n)$ decays exponentially in n , if L grows linearly in n .

Corollary 1. (Noise-induced barren plateaus). Let $N_{lm}, N_O \in \mathcal{O}(\text{poly}(n))$ and let $\eta_{lm}^i, \omega^j \in \mathcal{O}(\text{poly}(n))$ for all i, j . Then the upper bound $F(n)$ in Eq. (12) vanishes exponentially in n as

$$F(n) \in \mathcal{O}(2^{-an}), \quad (13)$$

for some positive constant a if we have

$$L \in \Omega(n). \quad (14)$$

The asymptotic scaling in Eq. (13) is independent of l and m , i.e., the scaling is blind to the layer, or the parameter within the layer, for which the derivative is taken. This corollary implies that when Eq. (14) holds, i.e. L grows at least linearly in n , the partial derivative $|\partial_{\theta_{lm}} \tilde{C}|$ exponentially vanishes in n across the entire cost landscape. In other words, one observes a Noise-Induced Barren Plateau (NIBP). We note that Eq. (14) is satisfied for all $q < 1$. That is, NIBPs occur regardless of the noise strength, it only changes the severity of the exponential scaling.

In addition, Corollary 1 implies that NIBPs are conceptually different from noise-free barren plateaus. First, NIBPs are independent of the parameter initialization strategy or the locality of the cost function. Second, NIBPs exhibit exponential decay of the gradient itself; not just of the variance of the gradient, which is the hallmark of noise-free barren plateaus. Noise-free barren plateaus allow the global minimum to sit inside deep, narrow valley in the landscape³⁴, whereas NIBPs flatten the entire landscape.

One of the strategies to avoid the noise-free barren plateaus is to correlate parameters, i.e., to make a subset of the parameters equal to each other³⁷. We generalize Theorem 1 in the following remark to accommodate such a setting, consequently showing that such correlated or degenerate parameters do not help in avoiding NIBPs. In this setting, the result we obtain in Eq. (16) below is essentially identical to that in Eq. (12) except with an additional factor quantifying the amount of degeneracy.

Remark 1. (Degenerate parameters). Consider the ansatz defined in Eqs. (1) and (3). Suppose there is a subset G_{st} of the set $\{\theta_{lm}\}$ in this ansatz such that G_{st} consists of g parameters that are degenerate:

$$G_{st} = \{\theta_{lm} \mid \theta_{lm} = \theta_{st}\}. \quad (15)$$

Here, θ_{st} denotes the parameter in G_{st} for which $N_{lm} \|\eta_{lm}\|_\infty$ takes the largest value in the set. (θ_{st} can also be thought of as a reference parameter to which all other parameters are set equal in value.) Then the partial derivative of the noisy cost with respect to θ_{st} is bounded as

$$|\partial_{\theta_{st}} \tilde{C}| \leq \sqrt{8 \ln 2} g N_O \|H_{lm}\|_\infty \|\omega\|_\infty n^{1/2} q^{L+1}, \quad (16)$$

at all points in the cost landscape.

Remark 1 is especially important in the context of the QAOA and the UCC ansatz, as discussed below. We note that, in the general case, a unitary of the form of Eq. (3) cannot be implemented as a single gate on a physical device. In practice one needs to compile the unitary into a sequence of native gates. Moreover, Hamiltonians with non-commuting terms are usually approximated with techniques such as Trotterization. This compilation overhead potentially leads to a sequence of gates that grows with n . Remark 1 enables us to account for such scenarios, and we elaborate on its relevance to specific applications in the next subsection.

In reality, noise on quantum hardware can be non-local. For instance in certain systems one can have cross-talk errors or coherent errors. We address such extensions to our noise model in the following remark.

Remark 2. (Extensions to the noise model). Consider a modification to each layer of noise \mathcal{N} in Eq. (8) to include additional k -local Pauli noise and correlated coherent (unitary) noise across multiple qubits. Under such extensions to the noise model, we obtain the same scaling results as those obtained in Lemma 1 and

Theorem 1. We discuss this in more detail in Supplementary Note 5.

Finally, we present an extension of our main result to the case of measurement noise. Consider a model of measurement noise where each local measurement independently has some bit-flip probability given by $(1 - q_M)/2$, which we assume to be symmetric with respect to the 0 and 1 outcomes. This leads to an additional reduction of our bounds on the cost function and its gradient that depends on the locality of the observable O .

Proposition 1. (Measurement noise). Consider expanding the observable O as a sum of Pauli strings, as in Eq. (4). Let w denote the minimum weight of these strings, where the weight is defined as the number of non-identity elements for a given string. In addition to the noise process considered in Fig. 2, suppose there is also measurement noise consisting of a tensor product of local bit-flip channels with bit-flip probability $(1 - q_M)/2$. Then we have

$$\left| \tilde{C} - \frac{1}{2^n} \text{Tr}[O] \right| \leq q_M^w G(n) \left\| \rho - \frac{1}{2^n} \right\|_1 \quad (17)$$

and

$$|\partial_{l_m} \tilde{C}| \leq q_M^w F(n) \quad (18)$$

where $G(n)$ and $F(n)$ are defined in Lemma 1 and Theorem 1, respectively.

Proposition 1 goes beyond the noise model considered in Theorem 1. It shows that in the presence of measurement noise there is an additional contribution from the locality of the measurement operator. It is interesting to draw a parallel between Proposition 1 and noise-free barren plateaus, which have been shown to be cost-function dependent and in particular depend on the locality of the observable O ³⁴. The bounds in Proposition 1 similarly depend on the locality of O . For example, when $w = n$, i.e., global observables, the factor q_M^w will hasten the exponential decay. On the other hand, when $w = 1$, i.e., local observables, the scaling is unaltered by measurement noise. In this sense, a global observable exacerbates the NIBP issue by making the decay more rapid with n .

Application-specific analytical results. Here we investigate the implications of our results from the previous subsection for two applications: optimization and chemistry. In particular, we derive explicit conditions for NIBPs for these applications. These conditions are derived in the setting where Trotterization is used, but other compilation strategies incur similar asymptotic behavior. We begin with the QAOA for optimization and then discuss the UCC ansatz for chemistry. Finally, we make a remark about the Hamiltonian Variational Ansatz (HVA), as well as remark that our results also apply to a generalized cost function that can employ training data.

Corollary 2. (Example: QAOA). Consider the QAOA with $2p$ trainable parameters, as defined in Eq. (5). Suppose that the implementation of unitaries corresponding to the problem Hamiltonian H_P and the mixer Hamiltonian H_M require k_P - and k_M -depth circuits, respectively. If local Pauli noise of the form in Eq. (7) with noise parameter q acts before and after each layer of native gates, then we have

$$|\partial_{\beta_l} \tilde{C}| \leq \sqrt{8 \ln 2} g_{l,P} N_P \|H_P\|_\infty \|\omega\|_\infty n^{1/2} q^{(k_P+k_M)p+1}, \quad (19)$$

$$|\partial_{\gamma_l} \tilde{C}| \leq \sqrt{8 \ln 2} g_{l,M} N_P \|H_M\|_\infty \|\omega\|_\infty n^{1/2} q^{(k_P+k_M)p+1}, \quad (20)$$

for any choice of parameters β_l, γ_l , and where $O = H_P$ in Eq. (2). Here $g_{l,P}$ and $g_{l,M}$ are the respective number of native gates parameterized by β_l and γ_l according to the compilation.

Corollary 2 follows from Remark 1 and it has interesting implications for the trainability of the QAOA. From Eqs. (19) and (20), NIBPs are guaranteed if pk_P scales linearly in n . This can manifest itself in a number of ways, which we explain below.

First, we look at the depth k_P required to implement one application of the problem unitary. Graph problems containing vertices of extensive degree such as the Sherrington-Kirkpatrick model inherently require $\Omega(n)$ depth circuits to implement⁵⁵. On the other hand, generic problems mapped to hardware topologies also have the potential to incur $\Omega(n)$ depth or greater in compilation cost. For instance, implementation of MaxCut and k -SAT using SWAP networks on circuits with 1-D connectivity requires depth $\Omega(n)$ and $\Omega(n^{k-1})$ respectively^{15,63}. Such mappings with the aforementioned compilation overhead for $k \geq 2$ are guaranteed to encounter NIBPs even for a fixed number of rounds p .

Second, it appears that p values that grow at least lightly with n may be needed for quantum advantage in certain optimization problems (for example⁶⁴⁻⁶⁷). In addition, there are problems employing the QAOA that explicitly require p scaling as $\text{poly}(n)$ ^{21,68}. Thus, without even considering the compilation overhead for the problem unitary, these QAOA problems may run into NIBPs particularly when aiming for quantum advantage. Moreover, weak growth of p with n combined with compilation overhead could still result in a NIBP.

Finally, we note that above we have assumed the contribution of k_P dominates that of k_M . However, it is possible that for choice of more exotic mixers¹⁶, k_M also needs to be carefully considered to avoid NIBPs.

Corollary 3. (Example: UCC). Let H denote a molecular Hamiltonian of a system of M_e electrons. Consider the UCC ansatz as defined in Eq. (6). If local Pauli noise of the form in Eq. (7) with noise parameter q acts before and after every $U_{l_m}(\theta_{l_m})$ in Eq. (6), then we have

$$|\partial_{\theta_{l_m}} \tilde{C}| \leq \sqrt{8 \ln 2} \hat{N}_{l_m} N_H \|\omega\|_\infty n^{1/2} q^{L+1}, \quad (21)$$

for any coupled cluster amplitude θ_{l_m} , and where $O = H$ in Eq. (2).

Corollary 3 allows us to make general statements about the trainability of UCC ansatz. We present the details for the standard UCC ansatz with single and double excitations from occupied to virtual orbitals^{50,69} (see Methods for more details). Let M_o denote the total number of spin orbitals. Then at least $n = M_o$ qubits are required to simulate such a system and the number of variational parameters grows as $\Omega(n^2 M_e^2)$ ^{63,70}. To implement the UCC ansatz on a quantum computer, the excitation operators are first mapped to Pauli operators using Jordan-Wigner or Bravyi-Kitaev mappings^{71,72}. Then, using first-order Trotterization and employing SWAP networks⁶³, the UCC ansatz can be implemented in $\Omega(n^2 M_e)$ depth, while assuming 1-D connectivity of qubits⁶³. Hence for the UCC ansatz, approximated by single- and double-excitation operators, the upper bound in Eq. (21) (asymptotically) vanishes exponentially in n .

To target strongly correlated states for molecular Hamiltonians, one can employ a UCC ansatz that includes additional, generalized excitations^{56,73}. A $\Omega(n^3)$ depth circuit is required to implement the first-order Trotterized form of this ansatz⁶³. Hence NIBPs become more prominent for generalized UCC ansatzes. Finally, we remark that a sparse version of the UCC ansatz can be implemented in $\Omega(n)$ depth⁶³. NIBPs still would occur for such ansatzes.

Additionally, we can make the following remark about the Hamiltonian Variational Ansatz (HVA). As argued in^{56,74,75}, the

HVA has the potential to be an effective ansatz for quantum many-body problems.

Remark 3. (Example: HVA). The HVA can be thought of as a generalization of the QAOA to more than two non-commuting Hamiltonians. It is remarked in ref. ⁵⁷ that for problems of interest the number of rounds p scales linearly in n . Thus, considering this growth of p and also the potential growth of the compiled unitaries with n , the HVA has the potential to encounter NIBPs, by the same arguments made above for the QAOA (e.g., Corollary 2).

Remark 4. (Quantum Machine Learning). Our results can be extended to generalized cost functions of the form $C_{\text{train}} = \sum_i \text{Tr}[O_i U(\theta) \rho_i U^\dagger(\theta)]$ where $\{O_i\}$ is a set of operators each of the form (4) and $\{\rho_i\}$ is a set of states. This can encapsulate certain quantum machine learning settings^{58–62} that employ training data $\{\rho_i\}$. As an example of an instance where NIBPs can occur, in one study⁶² an ansatz model has been proposed that requires at least linear circuit depth in n .

Numerical simulations of the QAOA. To illustrate the NIBP phenomenon beyond the conditions assumed in our analytical results, we numerically implement the QAOA to solve MaxCut combinatorial optimization problems. We employ a realistic noise model obtained from gate-set tomography on the IBM Ourense superconducting qubit device. In the Methods we provide additional details on the noise model and the optimization method employed.

Let us first recall that a MaxCut problem is specified by a graph $G = (V, E)$ of nodes V and edges E . The goal is to partition the nodes of G into two sets which maximize the number of edges connecting nodes between sets. Here, the QAOA problem Hamiltonian is given by

$$H_p = -\frac{1}{2} \sum_{ij \in E} C_{ij} (\mathbb{1} - Z_i Z_j), \quad (22)$$

where Z_i are local Pauli operators on qubit (node) i , $C_{ij} = 1$ if the nodes are connected and $C_{ij} = 0$ otherwise.

We analyze performance in two settings. First, we fix the problem size at $n = 5$ nodes (qubits) and vary the number of rounds p (Fig. 3). Second, we fix the number of rounds of QAOA at $p = 4$ and vary the problem size by increasing the number of nodes (Fig. 4).

In order to quantify performance for a given n and p , we randomly generate 100 graphs according to the Erdős-Rényi model⁷⁶, such that each graph G is chosen uniformly at random from the set of all graphs of n nodes. For each graph we run 10 instances of the parameter optimization, and we select the run that achieves the smallest energy. At each optimization step the cost is estimated with 1000 shots. Performance is quantified by the average approximation ratio when training the QAOA in the presence and absence of noise. The approximation ratio is defined as the lowest energy obtained via optimizing divided by the exact ground state energy of H_p .

In our first setting we observe in Fig. 3a that when training in the absence of noise, the approximation ratio increases with p . However, when training in the presence of noise the performance decreases for $p > 2$. This result is in accordance with Lemma 1, as the cost function value concentrates around $\text{Tr}[H_p]/2^n$ as p increases. This concentration phenomenon can also be seen clearly in Fig. 3b, where in fact we see evidence of exponential decay of cost value with p .

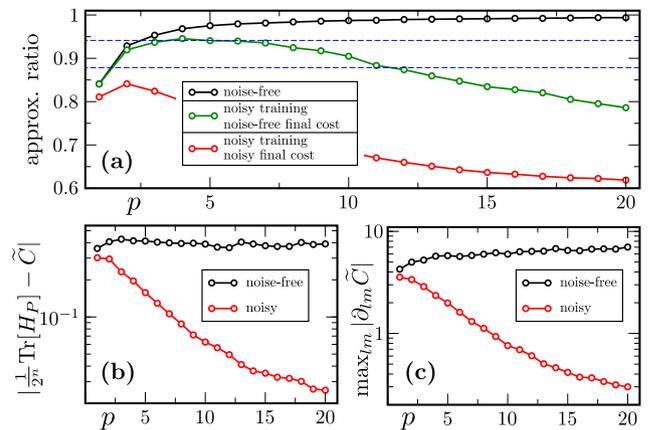


Fig. 3 QAOA heuristics in the presence of realistic hardware noise: increasing number of rounds for fixed problem size. **a** The approximation ratio averaged over 100 random graphs of 5 nodes is plotted versus number of rounds p . The black, green, and red curves respectively correspond to noise-free training, noisy training with noise-free final cost evaluation, and noisy training with noisy final cost evaluation. The performance of noise-free training increases with p , similar to the results in Ref. ¹⁵. The green curve shows that the training process itself is hindered by noise, with the performance decreasing steadily with p for $p > 4$. The dotted blue lines correspond to known lower and upper bounds on classical performance in polynomial time: respectively the performance guarantee of the Goemans-Williamson algorithm⁷⁷ and the boundary of known NP-hardness^{78,79}. **b** The deviation of the cost from $\text{Tr}[H_p]/2^n$ (averaged over graphs and parameter values) is plotted versus p . As p increases, this deviation decays approximately exponentially with p (linear on the log scale). **c** The absolute value of the largest partial derivative, averaged over graphs and parameter values, is plotted versus p . The partial derivatives decay approximately exponentially with p , showing evidence of Noise-Induced Barren Plateaus (NIBPs).

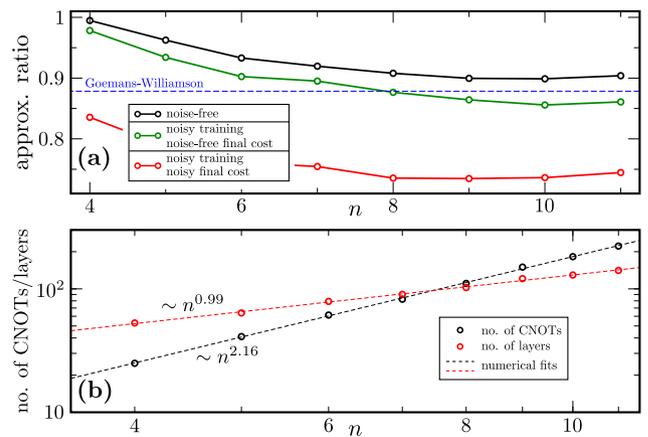


Fig. 4 QAOA heuristics in the presence of realistic hardware noise: increasing problem size for a fixed number of rounds. The approximation ratio averaged over 60 random graphs of increasing number of nodes n and fixed number of rounds $p = 4$ is plotted. The black, green, and red curves respectively correspond to noise-free training, noisy training with noise-free final cost evaluation, and noisy training with noisy final cost evaluation. **a** For a problem size of 8 nodes or larger, the noisily-trained approximation ratio falls below the performance guarantee of the classical Goemans-Williamson algorithm. **b** The depth of the circuit (red curve) scales linearly with the number of qubits, confirming we are in a regime where we would expect to observe Noise-Induced Barren Plateaus.

In addition, we can see the effect of NIBPs as Fig. 3a also depicts the value of the approximation ratio computed without noise by utilizing the parameters obtained via noisy training. Note that evaluating the cost in a noise-free setting has practical meaning, since the classicality of the Hamiltonian allows one to compute the cost on a (noise-free) classical computer, after training the parameters. For $p > 4$ this approximation ratio decreases, meaning that as p becomes larger it becomes increasingly hard to find a minimizing direction to navigate through the cost function landscape. Moreover, the effect of NIBPs is evident in Fig. 3c where we depict the average absolute value of the largest cost function partial derivative (i.e., $\max_{|m|} |\partial_m \tilde{C}|$). This plot shows an exponential decay of the partial derivative with p in accordance with Theorem 1.

Finally, in Fig. 3a we contextualize our results with previously known two-sided bounds on classical polynomial-time performance. The lower bound corresponds to the performance guarantee of the classical Goemans-Williamson algorithm⁷⁷, whilst the upper bound is at the value 16/17 which is the approximation ratio beyond which Max-Cut is known to be NP-hard^{78,79}.

In our second setting we find complementary results. In Fig. 4a we observe that at a problem size of 8 qubits or larger, 4 rounds of QAOA trained on the noisy circuit falls short of the performance guarantees of the classical Goemans-Williamson algorithm. As we increase the number of qubits, we also observe this increases the depth of the circuit linearly (Fig. 4b), thus confirming we are in a regime of NIBPs.

Our numerical results show that training the QAOA in the presence of a realistic noise model significantly affects the performance. The concentration of cost and the NIBP phenomenon are both also clearly visible in our data. Even though we observe performance for $n = 5$ and $p = 4$ that is NP-hard to achieve classically, any possible advantage would be lost for large problem sizes or circuit depth due to bad scaling. Hence, noise appears to be a crucial factor to account for when attempting to understand the performance of the QAOA.

Implementation of the HVA on superconducting hardware.

We further demonstrate the NIBP phenomenon in a realistic hardware setting by implementing the Hamiltonian Variational Ansatz (HVA) on the IBM Quantum *ibmq_montreal* 27-qubit superconducting device. At time of writing this holds the record for the largest quantum volume measured on an IBM Quantum device, which was demonstrated on a line of 6 qubits⁸⁰.

We implement the HVA for the Transverse Field Ising Model as considered in ref. ⁵⁷, with a local measurement $O = Z_0 Z_1$ on the first two qubits of the Ising chain. We assign the number of layers L of the ansatz to increase linearly with the number of qubits n according to the relationship $L = n - 1$. In order to minimize SWAP gates used in transpilation (and the accompanying noise that they incur), we modify each layer of the HVA ansatz to only include entangling gates between locally connected qubits.

Figure 5 plots the partial derivative of the cost function with respect to the parameter in the final layer of the ansatz, averaged over 100 random parameter sets. We also plot averaged cost differences from the corresponding maximally mixed values, as well as the variance of both quantities. In the noise-free case both the partial derivative and cost value differences decrease at a sub-exponential rate. Meanwhile, in the noisy case we observe that both the partial derivatives and cost value differences vanish exponentially until their variances reach the same order of magnitude as the shot noise floor. (As the shot budget on the IBM Quantum device is limited, this leads to a background of shot

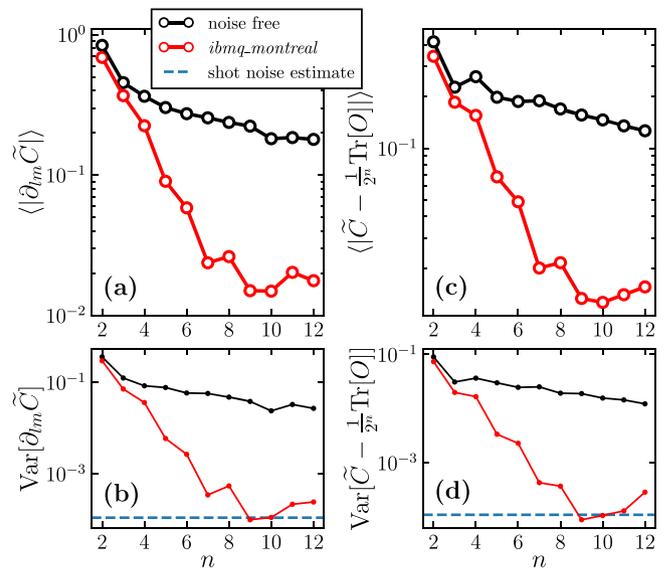


Fig. 5 Implementation on the *ibmq_montreal* superconducting-qubit device. We consider the HVA with the number of layers growing linearly in the number of qubits, n . **a** The average magnitude of the partial derivative of the noisy and noise-free cost, with respect to the parameter in the final layer, is plotted versus n . The average is taken over 100 randomly selected parameter sets. As n increases, the noisy average partial derivative decreases approximately exponentially, until around $n = 9$. This shows evidence of Noise Induced Barren Plateaus on real quantum hardware. **b** The deviation from exponential scaling can be understood by observing that it coincides with the point that the variance of the noisy partial derivatives reaches the same order of magnitude as the shot noise given by a finite sample budget of 8192 shots. Thus, from this point onward we expect fluctuations in the partial derivative to be dominated by shot noise, and gradients to be unresolvable. **c** The difference of the cost value from its corresponding maximally mixed value is plotted versus n . **d** The variance of this difference is plotted versus n . Both these quantities also show exponential decay until the variance of cost difference approaches the shot noise floor, which shows evidence of exponential cost concentration on this device.

noise, and we plot the order of magnitude of this with a dotted line.) This explicitly demonstrates that the problem of barren plateaus is one of *resolvability*. In principle, if one has access to exact cost values and gradients one may be able to navigate the cost landscape, however, the number of shots required to reach the necessary resolution increases exponentially with n .

Discussion

The success of NISQ computing largely depends on the scalability of Variational Quantum Algorithms (VQAs), which are widely viewed as the best hope for near-term quantum advantage for various applications. Only a small handful of works have analytically studied VQA scalability, and there is even less known about the impact of noise on their scaling. Our work represents a breakthrough in understanding the effect of local noise on VQA scalability. We rigorously prove two important and closely related phenomena: the exponential concentration of the cost function in Lemma 1 and the exponential vanishing of the gradient in Theorem 1. We refer to the latter as a Noise-Induced Barren Plateau (NIBP). Like noise-free barren plateaus, NIBPs require the precision and hence the algorithmic complexity to scale exponentially with the problem size. Thus, avoiding NIBPs is necessary for a VQA to have any hope of exponential quantum speedup.

NIBPs have conceptual differences from noise-free barren plateaus^{31–36} as the gradient vanishes with increasing problem size at every point on the cost function landscape, rather than probabilistically. As a consequence, NIBPs cannot be addressed by layer-wise training, correlating parameters and other strategies^{34,37,40–43}, all of which can help avoid noise-free barren plateaus. We explicitly demonstrate this in Remark 1 for the parameter correlation strategy. Similar to noise-free barren plateaus, NIBPs present a problem for trainability even when utilizing gradient-free optimizers³⁹ (e.g. simplex-based methods such as⁸¹ or methods designed specifically for quantum landscapes⁸²) or optimization strategies that use higher-order derivatives³⁸. At the moment, the only strategies we are aware of for avoiding NIBPs are: (1) reducing the hardware noise level, or (2) improving the design of variational ansatzes such that their circuit depth scales more weakly with n . Our work provides quantitative guidance for how to develop these strategies.

We emphasize that naïve mitigation strategies such as artificially increasing gradients cannot remove the exponential scaling of NIBPs as this simply increases the variance of any finite-shot evaluation of derivatives, and it does not improve the resolvability of the landscape. This argument extends simply to include any error mitigation strategy that implements an affine map to cost values^{83–89}. Further, most error mitigation techniques consist only of postprocessing noisy circuits. Thus, we deem it unlikely many strategies can remove exponential NIBP scaling as information about the cost landscape has fundamentally been lost (or at least been made exponentially inaccessible). This is in contrast to error correction where information is protected and recovered. However, in general it is an open question as to whether or not error mitigation strategies can mitigate NIBPs, and we leave this question for future work.

An elegant feature of our work is its generality, as our results apply to a wide range of VQAs and ansatzes. This includes the two most popular ansatzes, QAOA for optimization and UCC for chemistry, which Corollaries 2 and 3 treat respectively. In recent times QAQA, UCC, and other physically motivated ansatzes have been touted as the potential solution to trainability issues due to (noise-free) barren plateaus, while Hardware Efficient ansatzes, which minimize circuit depth, have been regarded as problematic. Our work swings the pendulum in the other direction: any additional circuit depth that an ansatz incorporates (regardless of whether it is physically motivated) will hurt trainability and potentially lead to a NIBP. This suggests that Hardware Efficient ansatzes are in fact worth exploring further, provided one has an appropriate strategy to avoid noise-free barren plateaus. This claim is supported by recent state-of-the-art implementations for optimization⁵⁵ and chemistry⁵⁴ using such ansatzes. Our work also provides additional motivation towards the pursuit of adaptive ansatzes^{90–98} that reduce circuit depth.

We believe our work has particular relevance to optimization. For combinatorial optimization problems, such as MaxCut on 3-regular graphs, the compilation of a single instance of the problem unitary e^{-iyH_p} can require an $\Omega(n)$ -depth circuit⁵⁵. Therefore, for a constant number of rounds p of the QAOA, the circuit depth grows at least linearly with n . From Theorem 1, it follows that NIBPs can occur for practical QAOA problems, even for constant number of rounds. Furthermore, even neglecting the aforementioned linear compilation overhead, NIBPs are guaranteed (asymptotically) if p grows in n . Such growth has been shown to be necessary in certain instances of MaxCut⁶⁴ as well as for other optimization problems^{21,68}, and hence NIBPs are especially relevant in these cases.

While it is well known that decoherence ultimately limits the depth of quantum circuits in the NISQ era, there was an interesting open question (prior to our work) as to whether one could

still train the parameters of a variational ansatz in the high decoherence limit. This question was especially important for VQAs for optimization, compiling, and linear systems, which are applications that do not require accurate estimation of cost functions on the quantum computer. Our work essentially provides a negative answer to this question. Naturally, important future work will involve extending our results to more general (e.g., non-unital) noise models, and numerically testing the tightness of our bounds. Moreover, our work emphasizes the importance of short-depth variational ansatzes. Hence a crucial research direction for the success of VQAs will be the development of methods to reduce ansatz depth.

Methods

Special cases of our ansatz. Here we discuss how the the QAOA, the Hardware Efficient ansatz, and the UCC ansatz fit into the framework as described in the general framework subsection.

1. Quantum Alternating Operator Ansatz. The QAOA can be understood as a discretized adiabatic transformation where the goal is to prepare the ground state of a given Hamiltonian H_p . The order p of the Trotterization determines the solution precision and the circuit depth. Given an initial state $|s\rangle$, usually the linear superposition of all elements of the computational basis $|s\rangle = |+\rangle^{\otimes n}$, the ansatz corresponds to the sequential application of two unitaries $U_p(\gamma) = e^{-iy_i H_p}$ and $U_M(\beta) = e^{-i\beta_j H_M}$. These alternating unitaries are usually known as the problem and mixer unitary, respectively. Here $\gamma = \{\gamma_k\}_{k=1}^L$ and $\beta = \{\beta_k\}_{k=1}^L$ are vectors of variational parameters which determine how long each unitary is applied and which must be optimized to minimize the cost function C , defined as the expectation value

$$C = \langle \gamma, \beta | H_p | \gamma, \beta \rangle = \text{Tr}[H_p | \gamma, \beta \rangle \langle \gamma, \beta |], \quad (23)$$

where $|\gamma, \beta\rangle = U(\gamma, \beta)|s\rangle$ is the QAOA variational state, and where $U(\gamma, \beta)$ is given by (5). In Fig. 6a we depict the circuit description of a QAOA ansatz for a specific Hamiltonian where $k_p = 6$.

2. Hardware Efficient Ansatz. The goal of the Hardware Efficient ansatz is to reduce the gate overhead (and hence the circuit depth) which arises when implementing a general unitary as in (3). Hence, when employing a specific quantum hardware the parametrized gates $e^{-i\theta_m H_m}$ and the unparametrized gates W_{lm} are taken from a gate alphabet composed of native gates to that hardware. Figure 6b shows an example of a Hardware Efficient ansatz where the gate alphabet is composed of rotations around the y axis and of CNOTs.

3. Unitary Coupled Cluster Ansatz. This ansatz is employed to estimate the ground state energy of the molecular Hamiltonian. In the second quantization, and within the Born-Oppenheimer approximation, the molecular Hamiltonian of a system of M_e electrons can be expressed as:

$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s$, where $\{a_i^\dagger\}$ ($\{a_i\}$) are Fermionic creation (annihilation) operators. Here, h_{pq} and h_{pqrs} respectively correspond to the so-

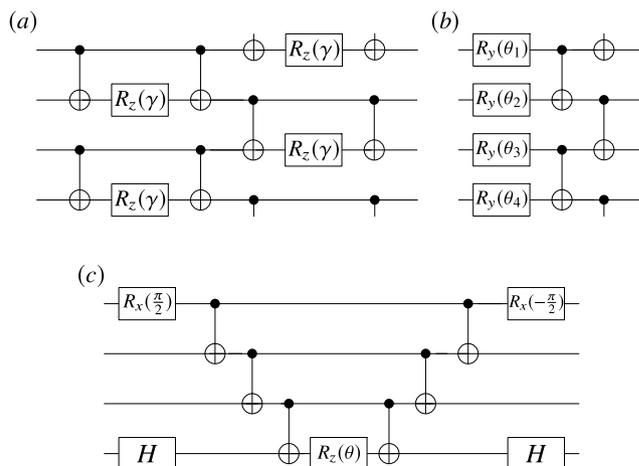


Fig. 6 Special cases of our general ansatz. **a** QAOA problem unitary e^{-iyH_p} for the ring-of-disagrees MaxCut problem, with Hamiltonian $H_p = \frac{1}{2} \sum_j Z_j Z_{j+1}$. **b** Hardware Efficient ansatz composed of CNOTs and single qubit rotations around the y -axis $R_y(\theta)$. **c** Unitary for the exponential $e^{-i\theta Y_1 Z_2 Z_3 X_4}$. This type of circuit is a representative component of the UCC ansatz.

called one- and two-electron integrals^{50,69}. The ground state energy of H can be estimated with the VQE algorithm by preparing a reference state, normally taken to be the Hartree-Fock (HF) mean-field state $|\psi_0\rangle$, and acting on it with a parametrized UCC ansatz.

The action of a UCC ansatz with single (T_1) and double (T_2) excitations is given by $|\psi\rangle = \exp(T - T^\dagger)|\psi_0\rangle$, where $T = T_1 + T_2$, and where

$$T_1 = \sum_{i \in \text{occ}} t_i^a a_i^\dagger a_i, \quad T_2 = \sum_{i,j \in \text{occ}} t_{ij}^{a,b} a_i^\dagger a_j^\dagger a_i a_j. \quad (24)$$

Here the i and j indices range over ‘‘occupied’’ orbitals whereas the a and b indices range over ‘‘virtual’’ orbitals^{50,69}. The coefficients t_i^a and $t_{ij}^{a,b}$ are called coupled cluster amplitudes. For simplicity, we denote these amplitudes $\{t_i^a, t_{ij}^{a,b}\}$ as $\{\theta_{lm}\}$. Similarly, by denoting the excitation operators $\{a_i^\dagger a_i, a_i^\dagger a_j^\dagger a_i a_j\}$ as $\{\tau_{lm}\}$, the UCC ansatz can be written in a compact form as $U(\theta) = e^{\sum_{lm} \theta_{lm} (\tau_{lm} - \tau_{lm}^\dagger)}$. In order to implement $U(\theta)$ one maps the fermionic operators to spin operators by means of the Jordan-Wigner or the Bravyi-Kitaev transformations^{71,72}, which allows us to write $(\tau_{lm} - \tau_{lm}^\dagger) = i \sum_{i,j} \mu_{lm}^i \sigma_i^x$. Then, from a first-order Trotterization we obtain (6). Here, $\mu_{lm}^i \in \{0, \pm 1\}$. In Fig. 6c we depict the circuit description of a representative component of the UCC ansatz.

Proof of Theorem 1. Here we outline the proof for our main result on Noise-Induced Barren Plateaus. We refer the reader to Supplementary Note 2 for additional details. We note that Lemma 1 and Remark 1 follow from similar steps and their proofs are detailed in Supplementary Notes 3 and 4 respectively. Moreover, we remark that Corollaries 1, 2 and 3 follow in a straightforward manner from a direct application of Theorem 1 and Remark 1.

Throughout our calculations we find it useful to use the expansion of operators in the Pauli tensor product basis. Given an n -qubit Hermitian operator Λ , one can always consider the decomposition

$$\Lambda = \lambda_0 \mathbb{1}^{\otimes n} + \lambda \cdot \sigma_n, \quad (25)$$

where $\lambda_0 \in \mathbb{R}$ and $\lambda \in \mathbb{R}^{4^n - 1}$. Note that here we redefine the vector of Pauli strings σ_n as a vector of length $4^n - 1$ which excludes $\mathbb{1}^{\otimes n}$.

Central to our proof is to understand how operators are mapped by concatenations of unitary transformations and noise channels. We do this through two lenses. First, given an operator Λ we investigate how various ℓ_p -norms of λ are related at different points in the evolution. Such quantities are well suited to study in our setting as we can use the transfer matrix formalism in the Pauli basis, that is, to represent a channel \mathcal{N} with the matrix $(T_{\mathcal{N}})_{ij} = \frac{1}{2^n} \text{Tr}[\sigma_i^j \mathcal{N}(\sigma_n^j)]$. Indeed, we see that the noise model in (7) has a diagonal Pauli transfer matrix, which motivates this choice of attack. The second quantity we use is the sandwiched 2-Rényi relative entropy $D_2(\rho \parallel \mathbb{1}^{\otimes n}/2^n)$ between a state ρ and the maximally mixed state. This is also useful to study due to the strong data processing inequality in ref. ⁹⁹ which quantifies how noise maps ρ closer to the maximally mixed state.

Let us now present two lemmas that reflect these two parts of the proof. The action of the noise in (7) on the operator Λ is to map the elements of λ as $\lambda_i \xrightarrow{\mathcal{N}} \lambda'_i = q_X^{x(i)} q_Y^{y(i)} q_Z^{z(i)} \lambda_i$ where $x(i)$, $y(i)$, and $z(i)$ respectively denote the number of X , Y , and Z operators in the i -th Pauli string. Recall the definition $q = \max\{|q_X|, |q_Y|, |q_Z|\}$. Since $x(i) + y(i) + z(i) \geq 1$, the inequality $|\lambda'| \leq q|\lambda|$ always holds. We use this relationship, along with Weyl’s inequality and the unitary invariance of Schatten norms to show that for an operator of the form (25) we have

$$\|\mathcal{W}^k(\Lambda)\|_\infty \leq \lambda_0 + q^k \|\vec{\lambda}\|_1 \quad (26)$$

where \mathcal{W}^k is a channel composed of k unitaries interleaved with noise channels of the form (7). The second lemma we present is a consequence of a strong data-processing inequality of the sandwiched 2-Rényi relative entropy of ref. ⁹⁹, from which we can show

$$D_2(\mathcal{W}^k(\rho) \parallel \mathbb{1}^{\otimes n}/2^n) \leq q^{2k} D_2(\rho \parallel \mathbb{1}^{\otimes n}/2^n) \quad (27)$$

where we note that $D_2(\rho \parallel \mathbb{1}^{\otimes n}/2^n)$ itself is always upper bounded by n for any n -qubit quantum state ρ .

Now that we have the main tools we present a sketch of the proof. In order to analyze the partial derivative of the cost function $\partial_{lm} \tilde{C} = \text{Tr}[O \partial_{lm} \rho_L]$ we first note that the output state ρ_L can be expressed as

$$\rho_L = (\mathcal{W}_a \circ \mathcal{W}_b)(\rho_0) = \mathcal{W}_a(\tilde{\rho}_1), \quad (28)$$

where ρ_0 is the input state and

$$\mathcal{W}_a = \mathcal{N} \circ \mathcal{U}_L \circ \dots \circ \mathcal{U}_{i+1} \circ \mathcal{N} \circ \mathcal{U}_{im}^+, \quad (29)$$

$$\mathcal{W}_b = \mathcal{U}_{im}^- \circ \mathcal{N} \circ \mathcal{U}_{l-1} \circ \dots \circ \mathcal{N} \circ \mathcal{U}_1 \circ \mathcal{N}, \quad (30)$$

where \mathcal{U}_{im}^\pm are channels that implement the unitaries $U_{im}^- = \prod_{1 \leq m \leq m} e^{-i\theta_m H_b}$ and $U_{im}^+ = \prod_{1 \leq m \leq m} e^{-i\theta_m H_b}$ such that $U_i = U_{im}^+ \cdot U_{im}^-$. For simplicity of notation here we have omitted the parameter dependence on the concatenation of channels.

Additionally, we have introduced the notation $\tilde{\rho}_1 = \mathcal{W}_b(\rho_0)$ and it is straightforward to show that

$$\partial_{lm} \tilde{\rho}_1 = -i[H_{lm}, \tilde{\rho}_1]. \quad (31)$$

Using the tracial matrix Hölder’s inequality¹⁰⁰, we can write

$$|\partial_{lm} \tilde{C}| = |\text{Tr}[\mathcal{W}_a^\dagger(O) \partial_{lm} \tilde{\rho}_1]| \quad (32)$$

$$\leq \|\mathcal{W}_a^\dagger(O)\|_\infty \|\partial_{lm} \tilde{\rho}_1\|_1, \quad (33)$$

where \mathcal{W}_a^\dagger is the adjoint map of \mathcal{W}_a . The two terms in the product can then be bounded with the above two techniques. Using (26) we find $\|\mathcal{W}_a^\dagger(O)\|_\infty \leq q^{L-l+1} N_O \|\omega\|_\infty$ for the first term. We bound the second term by using (31), a bound on Schatten norms of commutators¹⁰¹, quantum Pinsker’s inequality¹⁰², and (27) to obtain $\|\partial_{lm} \tilde{\rho}_1\|_1 \leq \sqrt{8 \ln 2} \|H_{lm}\|_\infty n^{1/2} q^l$. Putting the two parts together we obtain

$$|\partial_{lm} \tilde{C}| \leq \sqrt{8 \ln 2} N_O \|H_{lm}\|_\infty \|\omega\|_\infty n^{1/2} q^{L+1}, \quad (34)$$

completing the proof.

Proof of Proposition 1. Here we sketch the proof of Proposition 1, with additional details being presented in Supplementary Note 8.

We model measurement noise as a tensor product of independent local classical bit-flip channels, which mathematically corresponds to modifying the local POVM elements $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$ as follows:

$$P_0 = |0\rangle\langle 0| \rightarrow \tilde{P}_0 = \frac{1+q_M}{2} |0\rangle\langle 0| + \frac{1-q_M}{2} |1\rangle\langle 1| \quad (35)$$

$$P_1 = |1\rangle\langle 1| \rightarrow \tilde{P}_1 = \frac{1-q_M}{2} |0\rangle\langle 0| + \frac{1+q_M}{2} |1\rangle\langle 1|. \quad (36)$$

In turn, it follows that one can also model this measurement noise as a tensor product of local depolarizing channels with depolarizing probability $1 \geq (1 - q_M)/2 \geq 0$, which we indicate by \mathcal{N}_M . The channel is applied directly to the measurement operator such that $\mathcal{N}_M(O) = \sum_i \omega^i \mathcal{N}_M(\sigma_i^x) = \tilde{\omega} \cdot \sigma_n$. Here $\tilde{\omega}$ is a vector of coefficients $\tilde{\omega}^i = q_M^{w(i)} \omega^i$, where $w(i) = x(i) + y(i) + z(i)$ is the weight of the Pauli string. Here we recall that we have respectively defined $x(i)$, $y(i)$, $z(i)$ as the number of Pauli operators X , Y , and Z in the i -th Pauli string.

Let us first focus on the partial derivative of the cost. In the presence of measurement noise we then have

$$\partial_{lm} \tilde{C} = \frac{1}{2^n} \text{Tr}[(\tilde{\omega} \cdot \sigma_n)(g^{(L)} \cdot \sigma_n)] \quad (37)$$

$$= \tilde{\omega} \cdot g^{(L)}. \quad (38)$$

Which means that $|\partial_{lm} \tilde{C}| = |\tilde{\omega} \cdot g^{(L)}|$. We then examine the inner product in an element-wise fashion:

$$|\tilde{\omega} \cdot g^{(L)}| \leq \sum_i |\tilde{\omega}_i| |g_i^{(L)}| \leq \sum_i q_M^{w(i)} |\omega_i| |g_i^{(L)}|. \quad (39)$$

Therefore, defining $w = \min_i w(i)$ as the minimum weight of the Pauli strings in the decomposition of O , we have that $q_M^{w(i)} \leq q_M^w$, and hence we can replace $q_M^{w(i)}$ with q_M^w for each term in the sum. This gives an extra locality-dependent factor in the bound on the partial derivative:

$$|\partial_{lm} \tilde{C}| \leq q_M^w F(n). \quad (40)$$

An analogous reasoning leads to the following result for the concentration of the cost function:

$$\left| \tilde{C} - \frac{1}{2^n} \text{Tr} O \right| \leq q_M^w G(n). \quad (41)$$

Details of numerical implementations. The noise model employed in our numerical simulations was obtained by performing one- and two-qubit gate-set tomography^{103,104} on the five-qubit IBM Q Ourens superconducting qubit device. The process matrices for each gate native to the device’s alphabet, and the state preparation and measurement noise are described in ref. ⁹⁶ Appendix B]. In addition, the optimization for the MaxCut problems was performed using an optimizer based on the Nelder-Mead simplex method.

Data availability

Data generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Code availability

Code used for the current study is available from the corresponding author upon reasonable request.

Received: 4 February 2021; Accepted: 1 November 2021;

Published online: 29 November 2021

References

- Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
- Cerezo, M. et al. Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644 (2021).
- Endo, S., Cai, Z., Benjamin, S. C., & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. *Journal of the Physical Society of Japan* **90**, 032001 (2021).
- Bharti, K. et al. Noisy intermediate-scale quantum (nisq) algorithms. Preprint at <https://arxiv.org/abs/arXiv:2101.08448> (2021).
- Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).
- McClean, J. R., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *N. J. Phys.* **18**, 023023 (2016).
- Bauer, B., Wecker, D., Millis, A. J., Hastings, M. B. & Troyer, M. Hybrid quantum-classical approach to correlated materials. *Phys. Rev. X* **6**, 031045 (2016).
- Jones, T., Endo, S., McArdle, S., Yuan, X. & Benjamin, S. C. Variational quantum algorithms for discovering hamiltonian spectra. *Phys. Rev. A* **99**, 062304 (2019).
- Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
- Cirstoiu, C. et al. Variational fast forwarding for quantum simulation beyond the coherence time. *npj Quantum Inf.* **6**, 1–10 (2020).
- Heya, K., Nakanishi, K. M., Mitarai, K. & Fujii, K. Subspace variational quantum simulator. *Phys. Rev. Research* **1**, 033062 (2019).
- Yuan, X., Endo, S., Zhao, Q., Li, Y. & Benjamin, S. C. Theory of variational quantum simulation. *Quantum* **3**, 191 (2019).
- Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm. Preprint at <http://arxiv.org/abs/1411.4028> (2021).
- Wang, Z., Hadfield, S., Jiang, Z. & Rieffel, E. G. Quantum approximate optimization algorithm for MaxCut: a fermionic view. *Phys. Rev. A* **97**, 022304 (2018a).
- Crooks, G. E. Performance of the quantum approximate optimization algorithm on the maximum cut problem. Preprint at <http://arxiv.org/abs/1811.08419> (2021).
- Hadfield, S. et al. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms* **12**, 34 (2019).
- Bravo-Prieto, C. et al. Variational quantum linear solver: a hybrid algorithm for linear systems. Preprint at <https://arxiv.org/abs/1909.05820> (2019).
- Xu, X. et al. Variational algorithms for linear algebra. Preprint at <http://arxiv.org/abs/1909.03898> (2021).
- Koczor, B., Endo, S., Jones, T., Matsuzaki, Y. & Benjamin, S. C. Variational-state quantum metrology. *N. J. Phys.* <https://iopscience.iop.org/article/10.1088/1367-2630/ab965e> (2020).
- Meyer, J. J., Borregaard, J. & Eisert, J. A variational toolbox for quantum multi-parameter estimation. <https://arxiv.org/abs/2006.06303> (2020).
- Anschuetz, E., Olson, J., Aspuru-Guzik, A. & Cao, Y. Variational quantum factoring. In *Quantum Technology and Optimization Problems*. pp. 74–85 (Springer International Publishing, Cham, 2019) https://link.springer.com/chapter/10.1007/978-3-030-14082-3_7.
- Khatri, S. et al. Quantum-assisted quantum compiling. *Quantum* **3**, 140 (2019).
- Sharma, K., Khatri, S., Cerezo, M. & Coles, P. Noise resilience of variational quantum compiling. *N. J. Phys.* <https://iopscience.iop.org/article/10.1088/1367-2630/ab784c> (2020).
- Jones, T. & Benjamin, S. C. Quantum compilation and circuit optimisation via energy dissipation. <http://arxiv.org/abs/1811.03147>.
- Arrasmith, A., Cincio, L., Sornborger, A. T., Zurek, W. H. & Coles, P. J. Variational consistent histories as a hybrid algorithm for quantum foundations. *Nat. Commun.* **10**, 3438 (2019).
- Cerezo, M., Poremba, A., Cincio, L. & Coles, P. J. Variational quantum fidelity estimation. *Quantum* **4**, 248 (2020).
- Cerezo, M., Sharma, K., Arrasmith, A. & Coles, P. J. Variational quantum state eigensolver. Preprint at <https://arxiv.org/abs/2004.01372> (2020).
- LaRose, R., Tikku, A., O'Neel-Judy, E., Cincio, L. & Coles, P. J. Variational quantum state diagonalization. *npj Quantum Inf.* **5**, 1–10 (2019).
- Verdon, G., Marks, J., Nanda, S., Leichenauer, S. & Hidary, J. Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm. Preprint at <https://arxiv.org/abs/1910.02071> (2019).
- Johnson, P. D., Romero, J., Olson, J., Cao, Y. & Aspuru-Guzik, A. QVECTOR: an algorithm for device-tailored quantum error correction. Preprint at <https://arxiv.org/abs/1711.02249> (2017).
- McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 4812 (2018).
- Holmes, Z., Sharma, K., Cerezo, M. & Coles, P. J. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. Preprint at <https://arxiv.org/abs/arXiv:2101.02138> (2021).
- Sharma, K., Cerezo, M., Cincio, L. & Coles, P. J. Trainability of dissipative perceptron-based quantum neural networks. Preprint at <https://arxiv.org/abs/arXiv:2005.12458> (2020).
- Cerezo, M., Sone, A., Volkoff, T., Cincio, L. & Coles, P. J. Cost-function-dependent barren plateaus in shallow quantum neural networks. *Nature Communications* **12**, 1791 (2021).
- Marrero, C. O., Kieferová, M. & Wiebe, N. Entanglement induced barren plateaus. *PRX Quantum* **2**, 040316 (2021).
- Patti, T. L., Najafi, K., Gao, X. & Yelin, S. F. Entanglement devised barren plateau mitigation. *Phys. Rev. Research* **3**, 033090 (2021).
- Volkoff, T. & Coles, P. J. Large gradients via correlation in random parameterized quantum circuits. *Quantum Sci. Technol.* <http://iopscience.iop.org/article/10.1088/2058-9565/abd891> (2021).
- Cerezo, M. & Coles, P. J. Higher order derivatives of quantum neural networks with barren plateaus. *Quantum Sci. Technol.* **6**, 035006 (2021).
- Arrasmith, A., Cerezo, M., Czarnik, P., Cincio, L. & Coles, P. J. Effect of barren plateaus on gradient-free optimization. *Quantum* **5**, 558 (2021).
- Uvarov, A. & Biamonte, J. On barren plateaus and cost function locality in variational quantum algorithms. Preprint at <https://arxiv.org/abs/arXiv:2011.10530> (2020).
- Verdon, G. et al. Learning to learn with quantum neural networks via classical neural networks. Preprint at <https://arxiv.org/abs/arXiv:1907.05415> (2019).
- Grant, E., Wossnig, L., Ostaszewski, M. & Benedetti, M. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum* **3**, 214 (2019).
- Skolik, A., McClean, J. R., Mohseni, M., Smagt, P. & Leib, M. Layerwise learning for quantum neural networks. *Quantum Mach. Intell.* **3**, 5 (2021).
- Xue, C., Chen, Z.-Y., Wu, Y.-C. & Guo, G.-P. Effects of quantum noise on quantum approximate optimization algorithm. *Chinese Phys. Lett.* **38**, 030302 (2021).
- Marshall, J., Wudarski, F., Hadfield, S. & Hogg, T. Characterizing local noise in QAOA circuits. *IOP SciNotes* **1**, 025208 (2020).
- Gentini, L., Cuccoli, A., Pirandola, S., Verrucchi, P. & Banchi, L. Noise-resilient variational hybrid quantum-classical optimization. *Phys. Rev. A* **102**, 052414 (2020).
- Farhi, E. & Harrow, A. W. Quantum supremacy through the quantum approximate optimization algorithm. Preprint at <https://arxiv.org/abs/arXiv:1602.07674> (2016).
- Kübler, J. M., Arrasmith, A., Cincio, L. & Coles, P. J. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum* **4**, 263 (2020).
- Arrasmith, A., Cincio, L., Somma, R. D. & Coles, P. J. Operator sampling for shot-frugal optimization in variational algorithms. Preprint at <https://arxiv.org/abs/arXiv:2004.06252> (2020).
- Cao, Y. et al. Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**, 10856–10915 (2019).
- Bartlett, R. J. & Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **79**, 291 (2007).
- Lee, J., Huggins, W. J., Head-Gordon, M. & Whaley, K. B. Generalized unitary coupled cluster wave functions for quantum computation. *J. Chem. Theory Comput.* **15**, 311–324 (2018).
- Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242 (2017).
- Arute, F. et al. Hartree-Fock on a superconducting qubit quantum computer. *Science* **369**, 1084–1089 (2020a).
- Harrigan, M. P. et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nat. Phys.* **17**, 332–336 (2021).
- Wecker, D., Hastings, M. B. & Troyer, M. Progress towards practical quantum variational algorithms. *Phys. Rev. A* **92**, 042303 (2015).
- Wiersema, R. et al. Exploring entanglement and optimization within the hamiltonian variational ansatz. *PRX Quantum* **1**, 020319 (2020).
- Schuld, M., Sinayskiy, I. & Petruccione, F. The quest for a quantum neural network. *Quantum Inf. Process.* **13**, 2567–2586 (2014).
- Schuld, M., Sinayskiy, I. & Petruccione, F. An introduction to quantum machine learning. *Contemp. Phys.* **56**, 172–185 (2015).
- Biamonte, J. et al. Quantum machine learning. *Nature* **549**, 195–202 (2017).
- Beer, K. et al. Training deep quantum neural networks. *Nat. Commun.* **11**, 1–6 (2020).
- Abbas, A. et al. The power of quantum neural networks. *Nat. Comput. Sci.* **1**, 403–409 (2020).
- Gorman, B. O., Huggins, W. J., Rieffel, E. G. & Whaley, K. B. Generalized swap networks for near-term quantum computing. Preprint at <https://arxiv.org/abs/arXiv:1905.05118> (2019).
- Bravyi, S., Kliesch, A., Koenig, R. & Tang, E. Obstacles to Variational Quantum Optimization from Symmetry Protection. *Phys. Rev. Lett.* **125**, 260505 (2020).

65. Wang, Z., Hadfield, S., Jiang, Z. & Rieffel, E. G. Quantum approximate optimization algorithm for maxcut: a fermionic view. *Phys. Rev. A* **97**, 022304 (2018b).
66. Hastings, M. B. Classical and quantum bounded depth approximation algorithms. Preprint at <https://arxiv.org/abs/1905.07047> (2019).
67. Jiang, Z., Rieffel, E. G. & Wang, Z. Near-optimal quantum circuit for grover's unstructured search using a transverse field. *Phys. Rev. A* **95**, 062317 (2017).
68. Akshay, V., Philathong, H., Morales, M. E. S. & Biamonte, J. D. Reachability deficits in quantum approximate optimization. *Phys. Rev. Lett.* **124**, 090504 (2020).
69. McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. C. & Yuan, X. Quantum computational chemistry. *Rev. Mod. Phys.* **92**, 015003 (2020).
70. Romero, J. et al. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Sci. Technol.* **4**, 014008 (2018).
71. Ortiz, G., Gubernatis, J. E., Knill, E. & Laflamme, R. Quantum algorithms for fermionic simulations. *Phys. Rev. A* **64**, 022319 (2001).
72. Bravyi, S. B. & Kitaev, A. Y. Fermionic quantum computation. *Ann. Phys.* **298**, 210–226 (2002).
73. Noolij, M. Can the eigenstates of a many-body hamiltonian be represented exactly using a general two-body cluster expansion? *Phys. Rev. Lett.* **84**, 2108 (2000).
74. Ho, W. W. & Hsieh, T. H. Efficient variational simulation of non-trivial quantum states. *SciPost Phys.* **6**, 029 (2019).
75. Cade, C., Mineh, L., Montanaro, A. & Stanisic, S. Strategies for solving the fermi-hubbard model on near-term quantum computers. *Phys. Rev. B* **102**, 235122 (2020).
76. Erdos, P. & Renyi, A. On random graphs i. *Publ. math. Debr.* **6**, 18 (1959).
77. Goemans, M. X. & Williamson, D. P. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM* **42**, 1115–1145 (1995).
78. Arora, S., Lund, C., Motwani, R., Sudan, M. & Szegedy, M. Proof verification and the hardness of approximation problems. *J. ACM* **45**, 501–555 (1998).
79. Håstad, J. Some optimal inapproximability results. *J. ACM* **48**, 798–859 (2001).
80. Jurcevic, P. et al. Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Sci. Technol.* **6**, 025020 (2021).
81. Nelder, J. A. & Mead, R. A simplex method for function minimization. *Computer J.* **7**, 308–313 (1965).
82. Koczor, B. & Benjamin, S. C. Quantum analytic descent. Preprint at <https://arxiv.org/abs/2008.13774> (2020).
83. Czarnik, P., Arrasmith, A., Coles, P. J. & Cincio, L. Error mitigation with clifford quantum-circuit data. Preprint at <https://arxiv.org/abs/2005.10189> (2020).
84. Montanaro, A. & Stanisic, S. Error mitigation by training with fermionic linear optics. Preprint at <https://arxiv.org/abs/2102.02120> (2021).
85. Vovrosh, J. et al. Efficient mitigation of depolarizing errors in quantum simulations. Preprint at <https://arxiv.org/abs/2101.01690> (2021).
86. Rosenberg, E., Ginsparg, P. & McMahon, P. L. Experimental error mitigation using linear rescaling for variational quantum eigensolving with up to 20 qubits. Preprint at <https://arxiv.org/abs/2106.01264> (2021).
87. He, A., Nachman, B., de Jong, W. A. & Bauer, C. W. Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Phys. Rev. A* **102**, 012426 (2020).
88. Shaw, A. Classical-quantum noise mitigation for NISQ hardware. Preprint at <https://arxiv.org/abs/2105.08701> (2021).
89. Arute, F. et al. Observation of separated dynamics of charge and spin in the Fermi-Hubbard model. Preprint at <https://arxiv.org/abs/2010.07965> (2020).
90. Bilks, M., Cerezo, M., Verdon, G., Coles, P. J. & Cincio, L. A semi-agnostic ansatz with variable structure for quantum machine learning. Preprint at <https://arxiv.org/abs/2103.06712> (2021).
91. Grimsley, H. R., Economou, S. E., Barnes, E. & Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **10**, 1–9 (2019).
92. Tang, H. L. et al. qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor. *PRX Quantum* **2**, 020310 (2021).
93. Zhang, Z.-J., Kyaw, T. H., Kottmann, J., Degroote, M. & Aspuru-Guzik, A. Mutual information-assisted adaptive variational quantum eigensolver. *Quantum Sci. Technol.* **6**, 035001 (2021).
94. Rattew, A. G., Hu, S., Pistoia, M., Chen, R. & Wood, S. A domain-agnostic, noise-resistant, hardware-efficient evolutionary variational quantum eigensolver. Preprint at <https://arxiv.org/abs/1910.09694> (2019).
95. Chivilikhin, D. et al. MoG-VQE: Multiobjective genetic variational quantum eigensolver. Preprint at <https://arxiv.org/abs/2007.04424> (2020).
96. Cincio, L., Rudinger, K., Sarovar, M. & Coles, P. J. Machine learning of noise-resilient quantum circuits. *PRX Quantum* **2**, 010324 (2021).
97. Cincio, L., Subaşı, Y., Sornborger, A. T. & Coles, P. J. Learning the quantum algorithm for state overlap. *N. J. Phys.* **20**, 113022 (2018).
98. Du, Y., Huang, T., You, S., Hsieh, M.-H. & Tao, D. Quantum circuit architecture search: error mitigation and trainability enhancement for variational quantum solvers. Preprint at <https://arxiv.org/abs/2010.10217> (2020).
99. Hirche, C., Rouzé, C. & França, D. S. On contraction coefficients, partial orders and approximation of capacities for quantum channels. Preprint at <https://arxiv.org/abs/2011.05949> (2020).
100. Baumgartner, B. An inequality for the trace of matrix products, using absolute values. Preprint at <https://arxiv.org/abs/1106.6189> (2011).
101. Wenzel, D. & Audenaert, K. M. R. Impressions of convexity: an illustration for commutator bounds. *Linear algebra its Appl.* **433**, 1726–1759 (2010).
102. Ohya, M. & Petz, D. *Quantum entropy and its use* (Springer Science & Business Media, 2004) <https://www.springer.com/gp/book/9783540208068>.
103. Blume-Kohout, R. et al. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.* **8**, 1–13 (2017).
104. Nielsen, E. et al. Probing quantum processor performance with pyGSTi. *Quantum Sci. Technol.* **5**, 044002 (2020).
105. Müller-Hermes, A., França, D. S. & Wolf, M. M. Relative entropy convergence for depolarizing channels. *J. Math. Phys.* **57**, 022202 (2016).

Acknowledgements

We thank Daniel Stilck França for helpful discussions and for pointing us to ref. ¹⁰⁵. Research presented in this article was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20190065DR. SW and EF acknowledge support from the U.S. Department of Energy (DOE) through a quantum computing program sponsored by the LANL Information Science & Technology Institute. MC and AS were also supported by the Center for Nonlinear Studies at LANL. PJC also acknowledges support from the LANL ASC Beyond Moore's Law project. LC and PJC were also supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under the Quantum Computing Applications Team (QCAT) program.

Author contributions

The project was conceived by P.J.C., M.C., K.S. and L.C. Lemma 1 and Theorem 1 were proven by SW. Proposition 1 was proven by E.F. and M.C. Corollaries 2 and 3 were proven by K.S. and S.W. Numerical heuristics were run by L.C. Implementation on quantum hardware was run by S.W. The manuscript was written by S.W., E.F., K.S., M.C., A.S., L.C. and P.J.C.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27045-6>.

Correspondence and requests for materials should be addressed to Samson Wang, M. Cerezo or Patrick J. Coles.

Peer review information *Nature Communications* thanks Maria Kieferova and the other anonymous reviewer(s) for their contribution to the peer review this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021