Evaluating the noise resilience of variational quantum algorithms

Enrico Fontana,[1, 2, *] Nathan Fitzpatrick,[3] David Muñoz Ramo,[3] Ross Duncan,[3, 1, 4] and Ivan Rungger[2, †]

[1]*Department of Computer and Information Sciences,*
University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, UK
[2]*National Physical Laboratory, Teddington, TW11 0LW, United Kingdom*
[3]*Cambridge Quantum Computing Ltd, 9a Bridge Street, Cambridge, United Kingdom*
[4]*Department of Physics, University College London, Gower Street, London WC1E 6BT, UK*

We simulate the effects of different types of noise in state preparation circuits of variational quantum algorithms. We first use a variational quantum eigensolver to find the ground state of a Hamiltonian in presence of noise, and adopt two quality measures in addition to the energy, namely fidelity and concurrence. We then extend the task to the one of constructing, with a layered quantum circuit ansatz, a set of general random target states. We determine the optimal circuit depth for different types and levels of noise, and observe that the variational algorithms mitigate the effects of noise by adapting the optimized parameters. We find that the inclusion of redundant parameterised gates makes the quantum circuits more resilient to noise. For such overparameterised circuits different sets of parameters can result in the same final state in the noiseless case, which we denote as parameter degeneracy. Numerically, we show that this degeneracy can be lifted in the presence of noise, with some states being significantly more resilient to noise than others. We also show that the average deviation from the target state is linear in the noise level, as long as this is small compared to a circuit-dependent threshold. In this region the deviation is well described by a stochastic model. Above the threshold, the optimisation can converge to states with largely different physical properties from the true target state, so that for practical applications it is critical to ensure that noise levels are below this threshold.

I. INTRODUCTION

The rapid development of noisy intermediate-scale quantum (NISQ) computers [1] in recent years has seen the equally explosive rise of hybrid variational quantum algorithms (VQAs) [2–10]. VQAs are composed of a quantum subroutine embedded into a classical optimisation loop [2, 3, 11]. The quantum side of the algorithm consists of a quantum state preparation stage (or ansatz) with externally controllable parameters, and a measurement stage that returns the value of a cost function from the prepared quantum state. The classical loop is typically a gradient descent optimiser, which updates the parameters in order to minimise the cost function.

VQAs hold much promise for immediate application to NISQ era devices, not just because they do not require large qubit counts to be useful, but also because they are expected to offer some resilience to the noise that characterizes these devices [11–17]. It has been shown that VQAs can automatically compensate for coherent errors, such as over-/under-rotations [3, 14, 18]. This stems directly from the variational nature of VQAs, as errors that only shift the position of the cost function minimum do not affect the outcome of the optimisation [3]. The resilience of VQAs to decoherent noise has proved more difficult to characterize, although there is some theoretical evidence that they offer partial resilience to more general stochastic errors [14, 15]. Variational compiling has been shown to be robust against a noise model con-

taining gate and readout noise [13]. There further exist several methods for error mitigation, which allow to remove some unphysical states [15, 19–23]. However, it is widely believed that there exists a hard limit to the resilience to decoherent noise of VQAs. Recently it has been shown that noise in the state preparation circuit leads to an asymptotic flattening of the cost function landscape [24]. This phenomenon reflects an accumulation of decoherent errors that ultimately makes the optimization untrainable. The result is valid in the limit of deep quantum circuits, while there remains an intermediate regime in system size and circuit depth, in which VQAs can be run successfully. The understanding of the impact of noise in this intermediate regime remains lacking, and is the question that we address in this paper. We evaluate quantitatively how specific models of quantum noise impact the performance of simulated VQAs, and to what extent this can be mitigated by the circuit design and optimisation strategy.

The paper is divided into two main parts: the simulation of a variational quantum eigensolver (VQE) and a more general variational algorithm that aims at maximising fidelity with a random target state. For the former, we study the effect of varying the circuits and in particular the effect of adding redundant parameterised gates in presence of noise. As measures of quality we track the energy, fidelity and entanglement of the prepared state. For the latter, we employ a hardware-efficient ansatz [25], where the circuit is composed of a layer repeated several times. We present numerical results without and with noise-aware parameter optimisation, for various noise levels and number of layers in the circuit. We find that VQAs can partly mitigate for quantum noise in the state preparation circuit by optimizing the parameters in pres-

* enrico.fontana@npl.co.uk
† ivan.rungger@npl.co.uk

ence of noise, and demonstrate that the inclusion of redundant parameters, which in the noiseless case results in degenerate sets of parameters with equal energy, can lead to states with higher resilience to noise.

Finally, we find that for low noise levels the noise propagation in a VQA circuit targeting a random state is well described by a stochastic model. In this regime, the noisy state is a linearly modified form of the exact state. However, in some situations there exists a circuit-dependent noise threshold, above which the optimisation in presence of noise can lead to different states with potentially very different physical properties, such as their entanglement. For practical applications it is important to characterise this threshold in order to ensure that the noise is within the linear regime.

II. METHODS

A. Variational quantum algorithms

Mathematically, the ansatz of a VQA can be seen as a parameterised unitary operator $U(\boldsymbol{\theta})$, with the vector $\boldsymbol{\theta}$ representing the used parameters, that is applied to a fixed initial state ρ_0 to yield a desired output state $\rho(\boldsymbol{\theta}) = U(\boldsymbol{\theta})\rho_0 U(\boldsymbol{\theta})^\dagger$. Here we use density operators ρ to represent quantum states, as they can represent both pure and mixed states. Equivalently, we can associate to the unitary a superoperator $\mathcal{U}(\boldsymbol{\theta})\rho \coloneqq U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger$. In terms of superoperators the overall quantum circuit operation can be written as

$$\rho(\boldsymbol{\theta}) = \mathcal{U}(\boldsymbol{\theta})\,(\rho_0) = \mathcal{U}_L(\boldsymbol{\theta}_L)\mathcal{U}_{L-1}(\boldsymbol{\theta}_{L-1})\cdots\mathcal{U}_1(\boldsymbol{\theta}_1)\,(\rho_0),\tag{1}$$

where each unitary is parameterised by a set of parameters.

The cost function, $C(\boldsymbol{\theta})$, typically corresponds to the expectation value of an operator O for the prepared state. The classical optimiser then attempts to find the minimum of the real-valued cost function

$$C(\boldsymbol{\theta}) = \mathrm{Tr}[O\,\rho(\boldsymbol{\theta})].\tag{2}$$

Eqs. (1) and (2) encapsulate the unitary dynamics of a closed system representing a noiseless quantum computer. In our work we simulate this on a classical machine, which can be achieved by representing the operators in the equation as matrices and performing the operations numerically.

B. Noise model

A real quantum computer is an open quantum system, as it cannot be perfectly separated from its surroundings, and interactions with the environment cause deviations from unitarity, known as decoherence. Therefore, our basic unitary evolution needs to be expanded with a noise

model. We use the broad term quantum channel to describe both unitary and non-unitary dynamics, and when the channel is non-unitary it is termed a noisy quantum channel [26].

We construct the model by interleaving noiseless operations and noisy quantum channels that aim to replicate decoherent processes in quantum computers. Indicating the effect of noisy quantum channels as Λ_l, with the integer l indexing the specific channel, Eq. (1) is modified to

$$\rho(\boldsymbol{\theta}) = \Lambda_L \mathcal{U}_L(\boldsymbol{\theta}_L)\Lambda_{L-1}\cdots\Lambda_1\mathcal{U}_1(\boldsymbol{\theta}_1)\,(\rho_0).\tag{3}$$

We use identical noise channels throughout the circuit evaluation. Each of these spans all the qubits in the system, and is defined as the application of an identical one-qubit noise channel to every qubit:

$$\rho = \Lambda\,(\rho_{\mathrm{in}}) = \left(\bigotimes_{i=1}^{N}\Lambda^{(i)}\right)(\rho_{\mathrm{in}}),\tag{4}$$

where $\Lambda^{(i)}$ is the one-qubit channel acting on the ith qubit, ρ_{in} is the input state and N is the number of qubits. This is termed a product channel [27]. The same approximation is used in Ref. [28], and is valid when qubits are sufficiently separated physically and there is only small cross-talk. Note that for simplicity we have also assumed that the noise is identical on every qubit, however in general the qubits of a real quantum computer have noise characteristics that can differ significantly from one another [29]. In our approach we neglect cross-talk and coherent errors, as our main focus is on decoherent noise channels. Similarly, we do not consider readout noise, even though this is an important feature of real quantum devices[29]. The justification for this choice is that readout noise is independent from ansatz design, while the main question of this paper is to address the effect of a noisy ansatz itself. We also neglect finite sampling (shot) noise, which allows us to employ exact statevector simulations.

We consider three types of noise channels: amplitude damping, phase damping, and symmetric depolarising channels. Amplitude damping relates to the relaxation of the qubit from an excited state to its ground state, while phase damping to the loss of phase, perturbing the off-diagonal elements of the density matrix [30]. Two common metrics of qubit quality, the longitudinal relaxation ($T1$) and dephasing ($T2$) times, can be related directly to amplitude and phase damping [31–33]. Symmetric depolarising noise describes a decay to a completely mixed state, and hence is useful as a prototypical decoherent channel [30].

We make use of the Kraus operator formalism to apply these channels onto the quantum state, which carries the premise that quantum noise is a Markovian process [30]. In our implementation, the operators are parameterised by $\gamma \in [0,1]$, representing the strength of the noise. A value of 0 represents no noise (and hence an

identity channel), while a value of 1 is maximal in the sense that the output of the noise channel corresponds to the fixed point of the channel. For amplitude damping, this is the state $|0\rangle\langle 0|$, for dephasing it is any linear combination of $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, and for symmetric depolarising it is the completely mixed state. In Appendix A we provide the single-qubit Kraus operators used for the noise model, as well as details on how to obtain the product channel operators.

As a simplification we apply noise only on two-qubit gates. This is justified by physical considerations valid for most hardware systems, where 2-qubit operations are considerably slower than the 1-qubit ones, and have a much higher noise rate, usually by an order of magnitude in their gate fidelity figures [34]. Consistently with previous work [28, 35], we place a noisy channel after the gate, ensuring that at maximum noise the state output by the circuit will be unentangled even for non-depolarising noise channels like phase and amplitude damping.

C. Quality measures

We investigate two classes of variational quantum algorithms: VQE, where the energy is minimised, and target state optimisation, where the infidelity with respect to a random target state is minimised.

1. VQE simulations

As first system we examine the VQE, the prototypical example of VQA [2], and the results are presented in Sec. III. The VQE algorithm seeks to identify the ground state of a given Hamiltonian H. The cost function of the VQE is determined by setting $O = H$ in Eq. (2), which returns the energy of the trial state:

$$C(\boldsymbol{\theta}) = E(\boldsymbol{\theta}) := \mathrm{Tr}\left[H\rho(\boldsymbol{\theta})\right]. \tag{5}$$

Besides the energy of the state, to assess the quality of the output after convergence we also consider the fidelity with respect to the exact ground state of H. The fidelity between two general states ρ and σ is defined as [36]

$$F(\sigma, \rho) := \left(\mathrm{Tr}\sqrt{\sqrt{\rho}\,\sigma\,\sqrt{\rho}}\right)^2$$
$$= \mathrm{Tr}\left[\sigma\rho\right], \tag{6}$$

where the last equality is valid whenever one of the two states is pure. If we denote the exact ground state as $|\psi_{\mathrm{gs}}\rangle$, the fidelity for a parameterised ansatz circuit corresponds to

$$F(\boldsymbol{\theta}) := \mathrm{Tr}\left[|\psi_{\mathrm{gs}}\rangle\langle\psi_{\mathrm{gs}}|\,\rho(\boldsymbol{\theta})\right] = \langle\psi_{\mathrm{gs}}|\,\rho(\boldsymbol{\theta})\,|\psi_{\mathrm{gs}}\rangle, \tag{7}$$

where $\rho(\boldsymbol{\theta})$ is obtained with Eq. (3).

Finally, since in VQE we are almost exclusively interested in Hamiltonians with entangled ground states, we use entanglement as a further important test of the quality of the output state. If we have a 2-qubit system, there exists a broad selection of measures of bipartite entanglement. Since the noisy circuit produces mixed states, we choose one that is valid in this regime, namely the concurrence, $Q(\rho)$ [37]. The concurrence is frequently used in literature, as it is monotonically related to entanglement of formation, a meaningful measure of entanglement, while being easier to calculate in practice [38–40]. It has the closed form

$$Q(\rho) := \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4). \tag{8}$$

Here λ_i are the eigenvalues, in decreasing order, of the Hermitian matrix $\sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}}$ (note the similarity to Eq. (6)), where $\tilde{\rho}$ is the spin-flipped density matrix $(\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$, with $*$ indicating complex conjugation. $Q(\rho) = 0$ if and only if ρ is a linear combination of product states, and $Q(\rho) = 1$ if and only if ρ is a Bell state [41].

When working with larger qubit numbers we cannot directly apply concurrence. Out of the several possible measures of multi-qubit entanglement[42], we choose the maximum concurrence taken over all pairs of qubits in the system [43, 44].

2. Random target state fidelity optimisation

In the second part of the article we investigate how noise affects a variational algorithm for solving a more general task. The results are presented in Sec. IV. We consider random state fidelity optimisation, where, rather than choosing a specific Hamiltonian and evaluating its ground state, we select at random a wavefunction as target. We address the question of how closely a layered quantum circuit can approximate such a general target wavefunction under different noise regimes. The optimisation procedure is therefore modified to maximising the fidelity (see Eq. (6)) with a target state ρ_{T}. Equivalently, the problem can be formulated as a minimisation of the infidelity, defined as

$$R := 1 - F, \tag{9}$$

and hence the cost function is

$$C(\boldsymbol{\theta}) = R(\rho_{\mathrm{T}}, \rho(\boldsymbol{\theta})) := 1 - F(\rho_{\mathrm{T}}, \rho(\boldsymbol{\theta})). \tag{10}$$

We consider pure target states, so that the cost function given in Eq. (2) can be applied in this case by choosing $O = \mathbb{1} - \rho_{\mathrm{T}}$.

We then extend this to the case where one is provided with a set of n_{T} pure target states sampled from a uniform distribution. As figure of merit we use the *average optimal infidelity* over the set, which we define as

$$\bar{R} = \frac{1}{n_{\mathrm{T}}} \sum_{n=1}^{n_{\mathrm{T}}} \min_{\boldsymbol{\theta}} R(\rho_{\mathrm{T},n}, \rho(\boldsymbol{\theta})), \tag{11}$$

where $\rho_{T,n}$ is the target state with index n. The same measure has been used recently in Ref. [45]. For each target state, an optimisation procedure is run in presence of noise. $\bar{R} = 0$ would imply that the quantum circuit can represent any N-qubit state in the ensemble exactly. The addition of noise is expected to increase $\bar{R}$, as mixed states cannot have perfect overlap with pure states. As the distribution of target states, we consider the Haar distribution over real states, which is the unique uniform distribution over a space of pure quantum states [46].

We perform two types of numerical simulations: in the first, we optimise the circuit without noise to obtain the optimal ansatz parameters for the ideal case, but evaluate the infidelity using the noisy circuit ("non-reoptimised"); in the second type we start from the noiseless optimum, but then reoptimise the circuit by performing gradient descent with the noise channels in place ("reoptimised"). The non-reoptimised cost function provides an upper bound to the reoptimised cost function, and the two will be equal only if the location of the minimum is unaffected by noise.

Finally, in order to isolate the effect of noise from other contributions to the infidelity, we consider the average optimal relative infidelity, which we define as

$$\bar{R}_{\mathrm{rel}} := \bar{R} - \bar{R}_{\mathrm{id}}, \qquad (12)$$

where the subscript id indicates the infidelity evaluated in the ideal noiseless case.

III. VQE FOR ENERGY MINIMISATION

In this section we employ VQE for the task of identifying the ground state energy of a fermionic Hamiltonian, first for 2 and then for 4 qubits.

A. Two qubit system

We consider the following Hamiltonian on two qubits:

$$\hat{H} = \hat{\sigma}_z^1 \hat{\sigma}_z^2 + \hat{\sigma}_x^1 + \hat{\sigma}_x^2. \qquad (13)$$

This is an example of a transverse-field Ising Hamiltonian [47], and appears in this form in dynamical mean field theory (DMFT) simulations of the single-impurity Anderson model (SIAM) for its 2-electron ground state [48]. Since this minimal Hamiltonian has an entangled ground state, it forms an ideal starting point for the investigation of the effect of noise. As will be shown in the subsequent sections, the conclusions found here are applicable also to wavefunctions obtained with more complicated Hamiltonians. Furthermore, as the Hamiltonian is real, we can restrict our choice to just those circuits that always output a real wavefunction, enabling us to significantly reduce the number of parameters.

As ansätze we choose the three circuits shown in Fig. 1, each of which can be shown analytically to be able
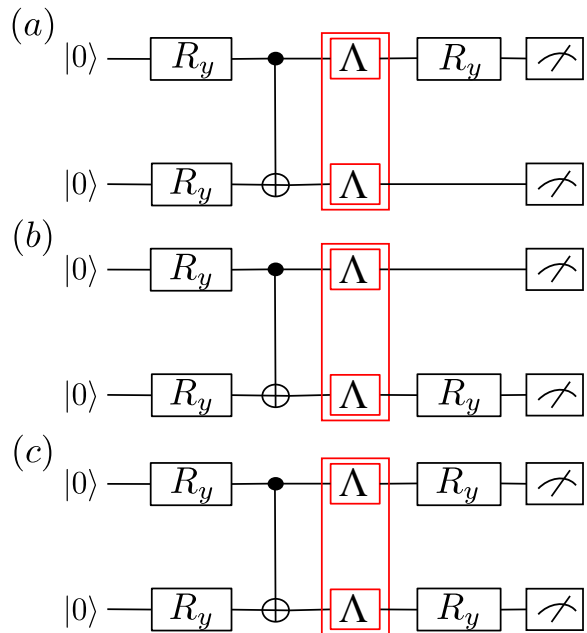

FIG. 1: Circuit ansätze used for state preparation on two qubits. Circuits (a) and (b) have 3 rotation parameters, and the circuit in (c) has 4 rotation parameters. In the noiseless case they all allow to cover the full real-states space of two qubits, and hence allow to construct any real 2-qubit state. There is one noise channel applied after the CNOT gate, as indicated by the Λ blocks on each qubit.

to prepare any possible real 2-qubit state. These ansätze include two inequivalent 3-parameter circuits, termed circuit (a) and (b), which differ on the position of the final rotation gate. We also consider a 4-parameter circuit that has rotations on both qubits before measurement. This latter circuit is over-parameterised, since it has one extra parameter compared to the previous two, and therefore allows us to explore the impact of redundant parameters.

The VQE algorithm is implemented using an exact density matrix simulator, which allows the use of a gradient-based classical optimiser, specifically the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimiser in our case [49–52]. The algorithm is evaluated on a range of $\gamma \in [0, 1]$, for phase, amplitude damping and symmetric depolarising noise, for all the ansatz circuits. As outlined in Sec. II B and Appendix A, γ is the parameter of the noise model that encodes the strength of the noise.

Analytically we calculate the exact ground state energy to be $E_{\mathrm{gs}} = -\sqrt{5} \approx -2.236$, and the concurrence to be $Q_{\mathrm{gs}} = \frac{1}{\sqrt{5}} \approx 0.447$. We verified that the numerical simulations with our used ansätze reproduce these analytical results exactly for $\gamma = 0$. The outcome of the noisy VQE simulation is shown in Fig. 2 for all state preparation circuits of Fig. 1. Straight away, we notice clear differences between the noise channels. In all measures of state quality, phase and amplitude damping channels are the least destructive, while the symmetric
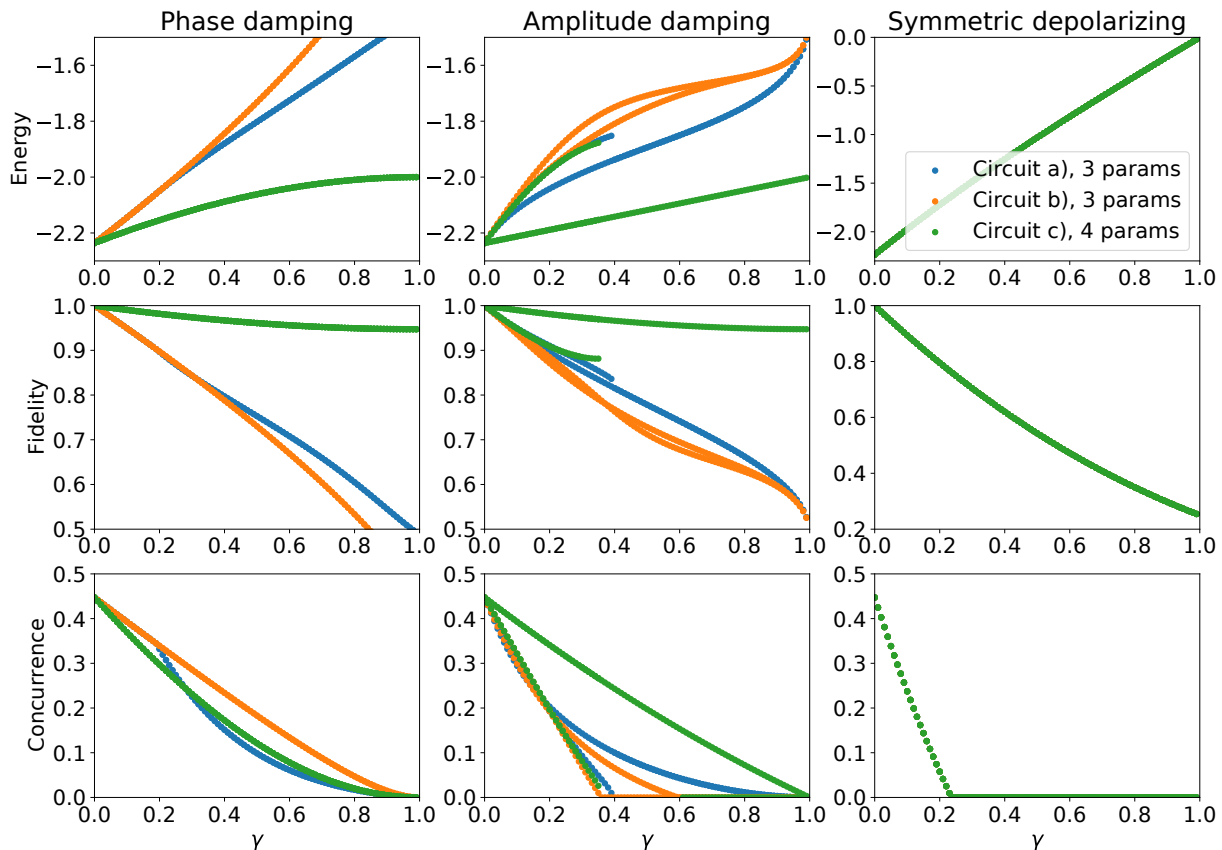
FIG. 2: Energy, fidelity and concurrence as function of the noise parameter γ for the all energy minima found for the 2-qubit circuits with 3 and 4 gates (Fig. 1), obtained applying the three different indicated noise types. We plot all the obtained local minima in the energy landscape, so that at a given value of γ there can be multiple points for the same circuit. For symmetric depolarising noise there is no circuit-dependent difference, and hence only one curve is visible as all curves overlap.

depolarising channel has a much more dramatic effect.

We perform a comprehensive search of the parameter space, allowing us to identify all the local minima in the energy landscape, which we represent by plotting multiple points for the same value of γ. The number of local minima shows a clear dependency on the noise type. Amplitude damping noise reveals local minima that branch out at low noise levels, resulting in multiple points for the same noise value. The number of minima depends on the circuit: although difficult to deduce from the figure, by numerically analyzing the results we find two solutions for circuits (a) and (b), and three solutions for the 4-parameter circuit. Symmetric depolarising and phase damping noise instead present a single global energy minimum.

Importantly, the extent to which different solutions are affected by noise depends strongly on the used circuit. While for symmetric depolarising noise there is no dependence on the circuit, for phase and amplitude damping noise the quality of the prepared state depends strongly on the used circuits. The fact that circuits (a) and (b) give different results shows that in presence of noise shifting a rotation gate from one qubit to another can improve

the quality. Furthermore, we consistently see that one of the solutions of the 4-parameter circuit is significantly better than any solution in the 3-parameter circuits, in all measures of state quality. This indicates that for this system the over-parameterised circuit with one redundant angle of rotation exhibits improved capabilities of noise mitigation, or equivalently a higher noise resilience.

B. Four qubit system

To determine how these findings generalize to higher qubit counts we perform the analogous analysis on a four-qubit system. We use the 2-local Hamiltonian

$$\hat{H} = \hat{\sigma}_z^1 \hat{\sigma}_z^3 + \frac{1}{2}(\hat{\sigma}_x^1 \hat{\sigma}_x^2 + \hat{\sigma}_y^1 \hat{\sigma}_y^2 + \hat{\sigma}_x^3 \hat{\sigma}_x^4 + \hat{\sigma}_y^3 \hat{\sigma}_y^4). \quad (14)$$

This Hamiltonian describes the same physical system as the one used for the previous section, but is valid for any number of electrons in the system [48]. For its ground state this Hamiltonian can be projected to the 2-qubit one discussed above, and it follows that the two Hamiltonians have the same ground state energy. As explained
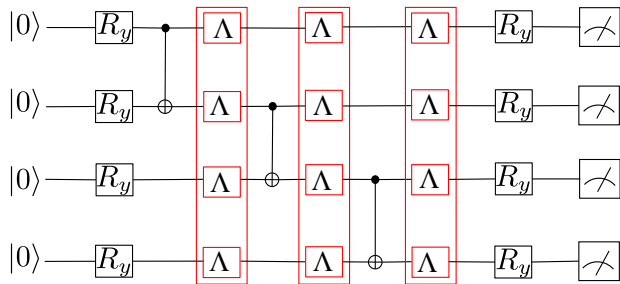
FIG. 3: Circuit used for the 4-qubit Hamiltonian VQE simulations. We apply a noise channel on all qubits after each CNOT gate.

in Section II C, since we are dealing with a system larger than 2 qubits we employ the maximum pairwise concurrence as a measure of entanglement. We find that the maximum pairwise concurrence of the new ground state is $Q_{\rm gs} = \frac{2}{\sqrt{5}} \approx 0.894$. The chosen state preparation ansatz circuit is shown in Fig. 3, and gives the exact energy in the noiseless case [48].

The results for maximum concurrence are shown alongside the results for energy and fidelity in Fig. 4. We see that the effect of the three types of noise channels on the state optimisation is very similar to the 2-qubit case. In particular, amplitude and phase damping have a less destructive impact than symmetric depolarising. Furthermore, the 4-qubit system exhibits multiple local minima at low noise for amplitude damping that are absent for phase and symmetric depolarising noise. However, compared to the 2-qubit system, the number of local minima is now larger and their appearance and disappearance more irregular.

C.   Discussion

Overall, the 2-qubit system and the 4-qubit system show a similar behavior with respect to the effect of the three types of noise channels, in all the measures of state quality. Consistently, the most destructive noise channel is the symmetric depolarising channel, where energy, fidelity and concurrence rapidly move away from the exact values. In both simulations the concurrence falls to zero at $\gamma \approx 0.2 - 0.3$.

In contrast, the other two noise channels concede a degree of robustness, yielding better performance even at high noise. Nevertheless, as noise rises towards $\gamma = 1$ all circuits eventually tend towards unentangled states. We note that for this system the concurrence is a much more stringent quality criterion than the energy and fidelity, since it inevitably goes towards zero for all circuits at $\gamma = 1$, while in some cases energy and fidelity only deviate by 10-20% from the exact value at maximum noise. For example, for 2-qubits and complete dephasing, the energy is only about 10% higher than the noiseless value. This shows that the energy alone can be a deceptively poor
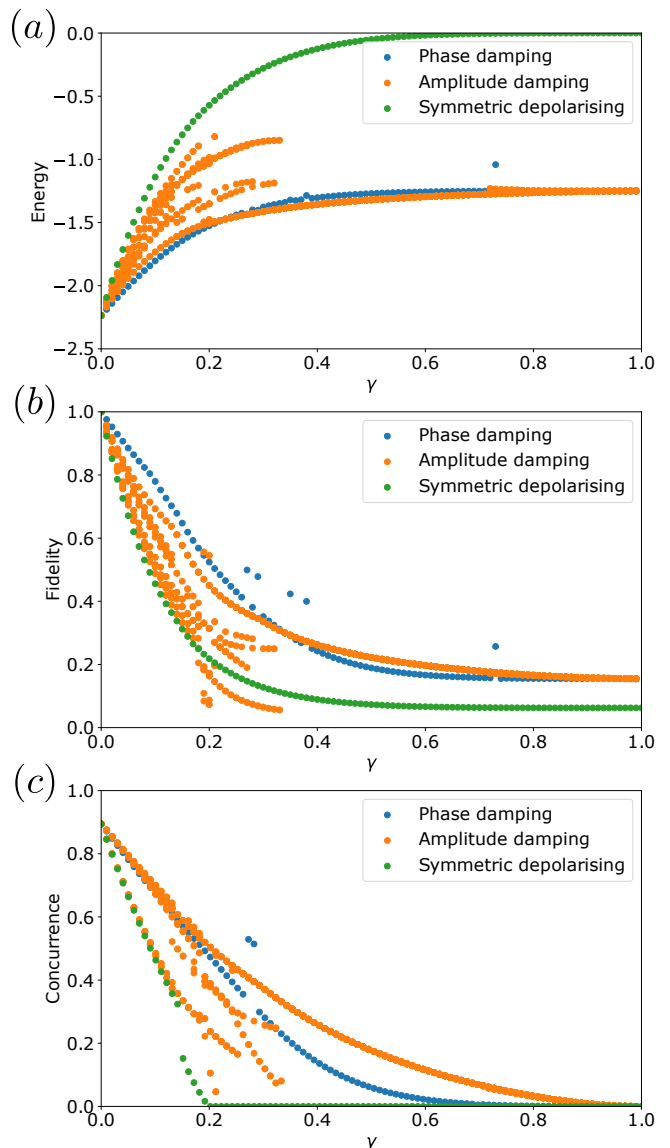

FIG. 4: Energy, fidelity and concurrence as function of the noise level γ for all the energy minima found for the 4-qubit VQE simulations with the circuit shown in Fig. 3, obtained applying the three different indicated noise types. We plot all the obtained local minima in the energy landscape, so that at a given value of γ there can be multiple points for the same circuit. Some points are missing due to imperfect optimisation, and due to the fact that local minima can appear and disappear for increasing noise levels.

quality measure for quantum algorithms.

Another common feature of the VQE experiments is the presence of multiple local minima when amplitude damping noise is present in the circuit, which are visualized as multiple lines branching off from the same point at zero noise as the noise parameter is increased. For some of these solutions the measures of quality decrease less with noise, and hence appear to be more resilient than others.

The damping noise models bring to light differences between otherwise equivalent circuits at zero noise. All circuits considered for the 2-qubit case (1) perform indistinguishably with respect to symmetric depolarising noise. Under the two damping noise channels, however, the three circuits are affected differently, with circuit (c) appearing to perform considerably better than circuits (a) and (b). Importantly, the different resilience to noise found for circuits (a) and (b) for the 2-qubit case shows that the circuit configuration needs to be optimised for maximisation of noise resilience. We expect that this optimization of gate placement is even more important when the noise level varies across qubits. Our results also show that the inclusion of redundant parameters can further improve the quality of the final state. The origin of this improvement is discussed in the next subsection.

D. Parameter degeneracies

In this section we show that the noise-induced phenomena discussed above, namely the presence of multiple local minima for amplitude damping and the differing performance of the three circuits with amplitude and phase damping, arise due to general features of parameterised quantum circuits. Let us consider those circuits for which there is a (non-identity) vector map f in parameter space such that:

$$U(\boldsymbol{\theta}) \left|\mathbf{0}\right\rangle = e^{i\phi}\, U(f(\boldsymbol{\theta})) \left|\mathbf{0}\right\rangle, \tag{15}$$

indicating equality up to an arbitrary global real phase ϕ. If such a map exists, we say that there is a *parameter degeneracy*, since two different sets of parameters result in the same identical state and hence energy (or more generally cost function). For a given circuit we can have a set of maps of similar form. In case the set is countable, we say that the circuit has *discrete parameter degeneracies*. Conversely, if the members of the set vary continuously across all the set, then the circuit has *continuous parameter degeneracies*.

A circuit with parameter degeneracies will feature symmetries of the cost function in the parameter space, since degenerate sets of parameters must yield the same value of the cost function. Therefore, the presence of parameter degeneracies implies the existence of multiple identical minima in the noiseless cost function. In the case of discrete degeneracies the minima are distinct and separated in the cost function landscape, while for a continuous degeneracy the minima are connected and may be visualized as a valley in the landscape. If one introduces a small state-dependent disturbance in the circuit, which breaks the symmetry between parameter degenerate states, such as for specific noise channels, this leads to different states being produced by the circuit. The degenerate global energy minima for the noiseless case then splits in local minima with different energies.

By analysing the parameters for the equivalent minima for the noiseless case in our 2- and 4-qubit VQE simu-
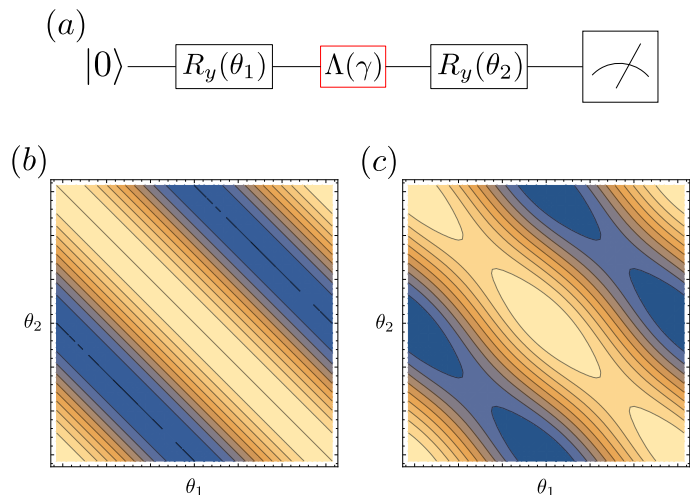

FIG. 5: Illustrative example for a continuous parameter degeneracy for an overparametrised circuit, where the second rotation by θ_2 is redundant in the absence of noise; the plots in (a) and (b) show the parameter space landscape of the cost function $C(\theta_1, \theta_2) = \text{Tr}[\rho(\theta_1, \theta_2) \left|0\right\rangle\langle 0|]$. In (a) the noiseless results are shown, and in (b) a phase damping channel is added between the Ry gates, with $\gamma = 0.4$. When noise is added, the continuous parameter degeneracy is broken, so that the valley in (a) is replaced by a set of minima in (b).

lations, we find that they obey fixed relations consisting of shifts of the angles by π and inversions. We verify that these relations are discrete parameter degeneracies by analytically showing that they preserve the state generated by the circuit. Furthermore, we find numerically that phase damping and symmetric depolarising channels preserve these degeneracies, while amplitude damping channels can break them. This is thus consistent with the multiple noise induced minima appearing for amplitude damping noise. In Ref. [53] an algorithm for the systematic construction of a particular set of discretely degenerate parameters is presented.

Continuous parameter degeneracies provide a useful framework for explaining the observed improved resilience to noise when including redundant rotation gates. Over-parameterised quantum circuits automatically have continuous parameter degeneracies, since any variation in the redundant parameter can be compensated by modifying the remaining parameters accordingly in order not to change the final state. As illustrative example circuit we consider the placement of two identical single-qubit rotation gates next to each other, shown in Fig. 5. As cost function we consider the overlap with the $\left|0\right\rangle$ state: $C(\theta_1, \theta_2) = \text{Tr}[\rho(\theta_1, \theta_2) \left|0\right\rangle\langle 0|]$. The computed cost function exhibits continuous parameter degeneracies in the noiseless case (Fig. 5(a)). When a phase damping noise channel is added between the two rotations, the continuous parameter degeneracies are broken, and there now exists a discrete set of minima (Fig. 5(b)). For such cases re-optimization of the parameters obtained for $\gamma = 0$ for the noisy case will generally lead to improved energies.
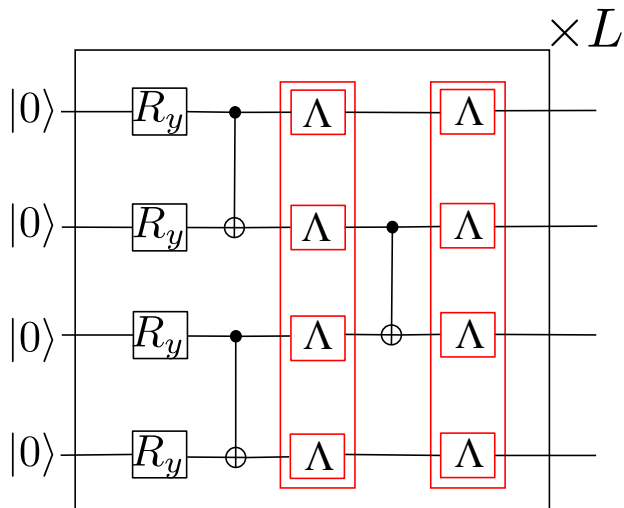
FIG. 6: Circuit block used for the random target state fidelity maximisation algorithm; shown is one layer of the ansatz, which is repeated a number L times in the full ansatz. Noise channels are added after CNOT gates.

Since at zero noise the minima exist as a continuous valley, the addition of noise introduces a gradient in this valley, which the optimizer can exploit to arrive to a better solution. The improved noise resilience for phase and amplitude damping noise due to a redundant parameter can therefore be explained as resulting from the ability of the circuit to explore more state preparation paths in parameter space when compared to a circuit with fewer parameters. It can therefore reach additional paths that are potentially less affected by the noise. For symmetric depolarising noise the relevant parameter degeneracies are not broken upon the addition of noise, and hence the addition of a redundant parameter does not improve the results.

IV. RANDOM TARGET STATES

To generalise the findings obtained for specific Hamiltonians, here we investigate the ability of a quantum circuit to represent an ensemble of random target states on 4 qubits. This approach allows to estimate how closely ground state wavefunctions of arbitrary Hamiltonians can be reproduced with a given circuit ansatz in presence of noise. For each target state we optimise the circuit to maximise the output state fidelity. We then average the maximum fidelity over the set of random states to obtain the average optimal infidelity (Eq. (11)). We consider again phase damping, amplitude damping and symmetric depolarising noise, and use the same density matrix simulator and local gradient-based minimiser (BFGS method) that we employed in the VQE simulations. The ensemble considered consists of 1000 real states generated by sampling a random orthogonal matrix from the circular real matrix distribution, the Haar distribution over real

orthogonal matrices [54], and picking its first column.

We choose a hardware-efficient ansatz [25] consisting of an identical layer of 4 rotation gates and 3 CX gates, repeated L times, as illustrated in Fig. 6. The choice of this specific structure is motivated by its high expressibility as demonstrated in Ref. [55], and by its compactness, which reduces the number of noise channels per layer. Indeed, the first two CX gates can be executed in parallel, and hence according to our noise model we insert only two noise channels per layer. Hardware-efficient circuits dense in parameterised operations are well-suited for preparing general quantum states [12, 19, 56].

Initially we consider the noise levels $\gamma \in \{0.1\%, 1\%, 5\%, 10\%\}$, which span the range of noise found in current devices [35]. The results are shown in Fig. 7. In the figure we present the average optimal fidelity for each γ for the three types of noise channels, where the vertical bars indicate the standard deviation over the ensemble. For comparison, we also plot the result for the noiseless case ($\gamma = 0$). We consider the two separate cases of noiseless training with noisy evaluation (non-reoptimised), and noisy training with noisy evaluation (reoptimised). By construction the reoptimised results are always better or equal to the non-reoptimised ones, as the former will take into account any noise-induced change in the cost function landscape. Comparing both results thus gives an insight on the degree to which noise affects the landscape, and on to what extent a variational algorithm can compensate for it. From Fig. 7 we can see that the reoptimised results in general improve significantly on the non-reoptimised ones.

We can observe that for $L \geq 4$ the noiseless fidelity is maximised for all the target states and equal to one, with zero standard deviation, showing that for such over-parametrised circuits any target state can be essentially exactly prepared. For $L < 4$ on the other hand the circuit does not reach all target states even without noise, and the standard deviation increases as L decreases. As expected, the addition of noise further reduces the average fidelity in all plots, with a larger noise level resulting in bigger deviation from the noiseless fidelity. For the non-reoptimised case, with all types of noise the fidelity reaches a peak in L, corresponding to the optimal circuit depth in presence of noise. Interestingly, in the case of phase and amplitude damping noise with reoptimisation, the fidelity continues to increase with the number of layers, even for high noise levels.

We also explore noise down to $\gamma = 10^{-4}$, which is representative of the higher quality end of current quantum devices [35]. Here we directly compare the relative infidelity as a measure of the effects of noise only, as a function of layers. The results are shown in Fig. 8 for phase damping noise (the results for other types of noise are shown in Appendix C). The relative infidelity reaches a peak at $L = 4$ for the reoptimised case, while for the non-reoptimised case it increases monotonically.

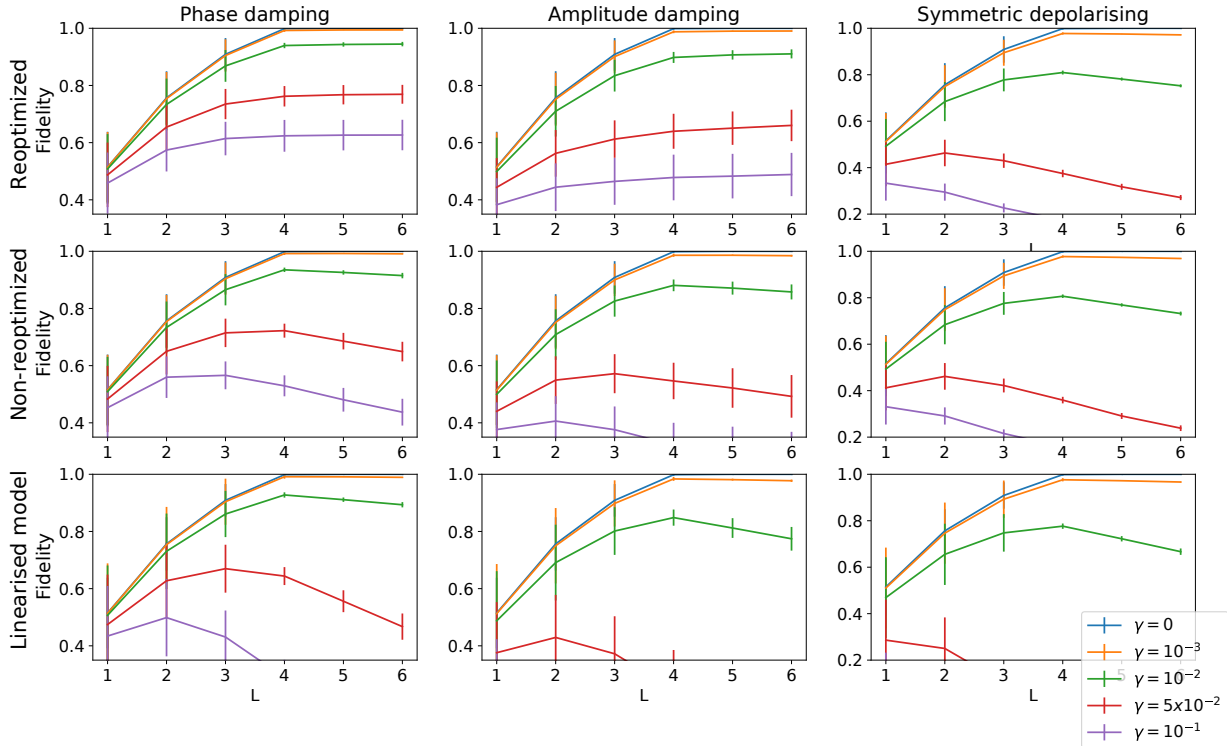For larger values of L we have an overparametrised

FIG. 7: Fidelity vs number of layers, L, at realistic noise levels, for the circuit ansätze illustrated in Fig. 6 ($\gamma = 0\%$ (blue), 0.1% (orange), 1% (green), 5% (red), 10% (purple)). The plots in the first column are for phase damping noise, the second column plots are for amplitude damping noise, and the third column plots are for symmetric depolarising noise. The top row of plots are for non-reoptimised parameters, in the second row of plots the rotation parameters are reoptimised at each noise level, and the in the third row of plots the linear noise model results are presented. Each point shows the average over 1000 target states, and the vertical bars at each point indicate the standard deviation.
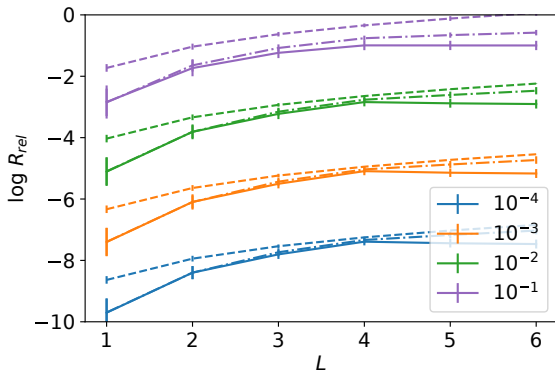

FIG. 8: Relative infidelity as function of number of layers, L, for different levels γ of phase damping noise. The solid curves are for noise-aware reoptimised parameters, the dash-dotted curves are for parameters fixed at their values optimized in absence of noise (non-reoptimised), and the dashed curves indicate the results of the linear noise model. Each point shows the average over 1000 target states, and the vertical bars at each point indicate the standard deviation.

ciruit, and we therefore expect the presence of a correspondingly large number of parameter degeneracies. For the 4-qubit VQE system and amplitude damping noise we found a large number of local minima due to discrete parameter degeneracies (Fig. 4). To show this is also the case for general target states we construct a set of degenerate states using the algorithm outlined in Ref. [53], and evaluate the fidelity for each of these sets of parameters. In Fig. 9 we show the resulting distribution of fidelities for $L = 4$ as example. While at zero noise they all give the same fidelity, with amplitude damping noise a rather large spread of the fidelities is found. For phase damping and symmetric depolarising noise on the other hand the degeneracies are preserved, and a single sharp peak is found in the histogram.

In the VQE optimisation (Section III) we observed that under symmetric depolarising and amplitude damping noise, there exists a threshold $\gamma < 1$, past which the algorithm converges to a non-entangled state, corresponding to an undesired noise-induced transition. It is important to verify whether a similar phenomenon appears for the fidelity maximisation with general target
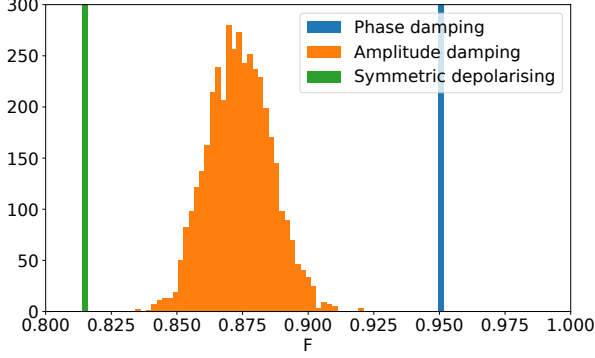
FIG. 9: Distribution of fidelity maxima originating from broken parameter degeneracies due to phase damping (blue), amplitude damping (orange), and symmetric depolarising (green) noise. The horizontal axis shows the fidelity, and the vertical axis the number of occurrences in a range of 0.002 around a given fidelity value. The total number of considered discrete parameter degenerate states is 4096, and is constructed using the algorithm described in Ref. [53]. The vertical axis has been cropped for phase damping and symmetric depolarising noise, where all states are within one bin. We choose a random real target state at $L = 4$ and $\gamma = 0.01$.

states. We therefore choose a target state at random from the distribution, and plot the fidelity of the output state after optimization as a function of noise. The results are shown in Fig. 10 for phase damping noise, $\gamma \in [0, 0.1]$ and $L = 3$. In Fig. 10a we show fidelity and concurrence, and in Fig. 10c we show a representative subset of the parameters optimized at each γ. In Fig. 10b we show fidelity and concurrence evaluated without noise, but with the circuit rotation parameters optimized with noise (Fig. 10c). We indeed observe a noise-induced transition in the example considered, which appears at a much lower noise level threshold ($\gamma \approx 0.04$) than in the VQE simulations. This transition is also visible in the converged parameter values (Fig. 10c) and the resulting quality measures evaluated for those parameters without noise (Fig. 10c). However, the state after the transition is still entangled, and indeed the concurrence behaves in a nontrivial way, highlighting that the transition is more complex in the general case. We note that the detailed behavior depends on the specific target state, and other examples of such transitions display different behaviours, ranging from sharp thresholds to smoother transitions (see Appendix C). There are also target states for which no well-defined transition can be observed.

A. Discussion

For the non-reoptimised case, the fidelity reaches a peak in L, which depends on the noise level, signaling the point where the noise from the increased number of
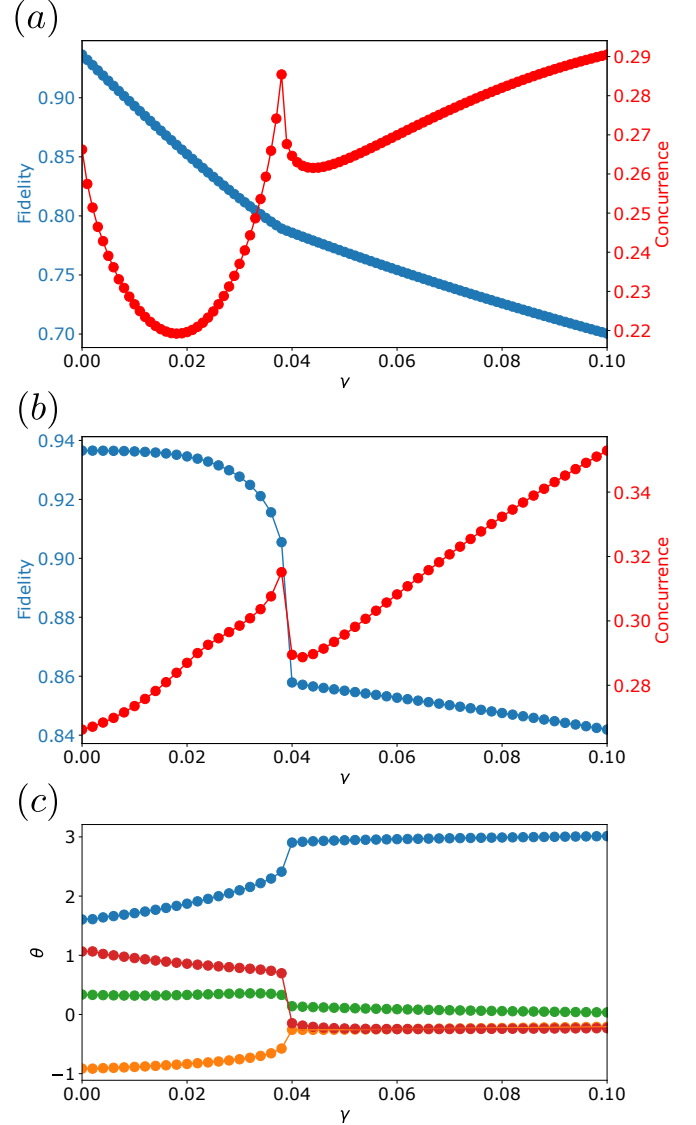

FIG. 10: Evolution of fidelity and concurrence at convergence for fidelity optimization at different strengths γ of phase damping noise, for a single random target state at $L = 3$. Four indicative optimised rotation angles are shown in the bottom panel to illustrate the evolution of the optimal parameters with increasing noise. The resulting fidelity and maximum pairwise concurrence are shown in the top panel. The central panel shows the measures evaluated without noise for the angles optimised at each γ value. A discontinuity in the slope of both state quality measures and converged angles is found at $\gamma \approx 0.04$, which therefore corresponds to the threshold γ value for this state and circuit, above which linear extrapolation of the properties to zero noise is not possible.

noise channels overcomes the improvement in the accuracy of the circuit with the additional parameters. Interestingly, in the case of phase and amplitude damping noise with reoptimisation the fidelity continues to increase with the number of layers even for high noise levels. The standard deviation of the results shows that

the most variability in the state quality is present for shallower circuits, and generally decreases with a larger number of layers. This suggests that, while the accuracy of shallow circuits depends heavily the target state, overparametrised circuits with more layers become increasingly consistent at approximating a general state. Interestingly, the pattern still holds in the presence of noise, however there is additional variance, showing the state-dependence of the noise effects. Past $L = 4$ the reoptimised simulations show a noticeable improvement in noisy fidelity compared to the non-reoptimised simulations. This depth threshold is significant, as it marks the point past which the circuit can perfectly reproduce all target states at zero noise. Any additional layer beyond $L = 4$ therefore does not contribute to the noiseless fidelity and only introduces redundancy. In the noisy case this overparametrisation leads to improved resilience for amplitude and phase damping noise, while for symmetric depolarising noise no improvement is found. This is analogous to what found for the VQE simulations in Sec. III. The specificity to amplitude and phase damping noise is due to the fact that these break the continuous degeneracies resulting from an over-parameterisation of the quantum circuit, while symmetric depolarising noise does not. Note that we do not expect the improvement to continue to arbitrarily large number of layers, as presumably the capability of over-parametrisation to minimise noise is bounded. More research on real hardware is needed to understand the practical limits.

Improved state quality upon parameter reoptimisation in quantum algorithms has previously been reported in Refs. [32] and [15], where the authors find that phase and amplitude damping noise generally impact state preparation in VQE less than symmetric depolarising noise, with phase damping noise being the least impactful. In particular, Ref. [15] studies how reoptimisation under noise significantly improves the results of a VQE for a chemical problem. However, their model applies noise after the state has been prepared, as opposed to the interleaved noise model proposed here. Hence, the breaking of parameter degeneracies could not be observed in that context.

1. Stochastic model

To provide further insight in our numerical results, here we present a model for the infidelity at small γ. In Appendix B we formulate a model that approximates noise propagation as linear, in the sense that each noise channel contributes an additive factor to the final relative infidelity. The model gives an estimate for the average relative infidelity and its variance as

$$\bar{R}_{\rm rel} \approx \alpha \, \gamma \, d, \tag{16}$$
$$\Delta_{\rm rel}^2 \approx \beta \, \gamma^2 d^2, \tag{17}$$

where α and β are constants obtained from the target state distribution and the noise channel, and d is the number of noise channels, which for our ansatz if $d = 2L$. In Appendix B we provide a detailed description including the procedure for the calculation of α and β, together with their numerical values obtained for our systems.

We plot the expectations of the model for the fidelity in Fig. 7 (bottom row of panels), and for the relative infidelity in Fig. 8. Overall the model captures the numerical trends rather well. As expected from the model being a linear approximation, it describes better the behavior of the fidelity at low noise levels, while for higher noise levels the deviations compared to the numerical results become larger. Since the model is formulated under the assumption of no noisy optimisation, it matches the non-reoptimised results much better than the reoptimised ones. In the latter case, the agreement is nevertheless reasonable until $L < 4$, however it diverges significantly for larger L, since the model does not include the improvement of fidelity due to parameter reoptimisation with the number of layers for phase and amplitude damping noise.

A further effect of the noisy optimisation that the linear model cannot capture is the sudden transition to more noise-resilient set of parameters above a noise threshold, as observed in Fig. 10. Clearly, this poses an upper bound to the noise level that such simplified models can adequately describe, as there exists a threshold noise level past which the new state cannot be extrapolated back to the noiseless state. For noise mitigation techniques that extrapolate finite noise data down to the zero noise level[20, 22], this implies that noise data needs to be collected below this critical threshold to avoid extrapolating to potentially undesired zero noise states.

V. CONCLUSIONS

We study the effects of different types and levels of noise on the quality of the results of VQAs. We find symmetric depolarising noise to be the most detrimental, while for amplitude and phase damping noise it is possible to mitigate the effects of noise by optimized gate placement, overparametrisation and noise-aware reoptimization. We obtain these results consistently across our considered systems, a 2- and 4-qubit simulation for a specific Hamiltonian, and 4-qubit simulations for general target states. We introduce the concept of parameter degeneracies, which are sets of parameters in the quantum circuit ansatz that give the same identical output state in the noiseless case. When noise is added, these degeneracies can be broken, leading to some of the originally degenerate parameter states to be more resilient to noise than others. For the VQE simulations we use three measures of state quality, namely energy, fidelity and entanglement, and show that the energy alone can be a deceptively poor quality measure for quantum algorithms. When maximising the fidelity of the state produced with

a given ansatz with a target state, we find that in presence of symmetric depolarising noise there is a circuit depth, where the fidelity is maximised. For amplitude and phase damping noise, and for the considered circuit depths, noise-aware parameter reoptimisation allows to progressively improve the fidelity as the circuit depth is increased. The results without noise-aware optimization compare well with a linearised noise model. We show that the average deviation from the target state is linear for low enough noise levels. For a number of target states and circuits there is a noise threshold, above which the states produced by the circuit can have largely different physical properties from the true target state. For practical applications it is critical to ensure that noise levels are below this threshold in order to preclude convergence to unphysical solutions.

VI.   ACKNOWLEDGEMENTS

APPENDIX

APPENDIX A: Kraus operators for simulations of noise

In the Kraus operator formalism, each noise channel is assigned a set of operators $\{E_i\}$, which are applied to a quantum state by conjugation. The integer index i spans across all considered operators for a given channel. For a given input quantum state ρ_{in} the noisy output state $\rho = \Lambda(\rho_{\text{in}})$ is obtained by

$$\rho = \Lambda(\rho_{\text{in}}) = \sum_k E_k \rho_{\text{in}} E_k^\dagger. \tag{A1}$$

In order to preserve the trace of the quantum state, Kraus operators need to obey the condition $\sum_k E_k^\dagger E_k = 1$.

The one-qubit phase damping channel is implemented with the following Kraus operator matrices:

$$E_1^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \ E_2^{(1)} = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\gamma} \end{pmatrix}. \tag{A2}$$

This channel has the effect of suppressing the off-diagonal components of the density matrix, while keeping the diagonal components unchanged. The one-qubit amplitude damping channel is described by:

$$E_1^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \ E_2^{(1)} = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}. \tag{A3}$$

Similarly to the phase channel, the off-diagonal components are suppressed. At the same time, however, the diagonal components are altered in favour of the $|0\rangle$ state, representing a relaxation of the system towards the ground state.

Finally, the operators for a single-qubit symmetric depolarising channel are:

$$E_1^{(1)} = \sqrt{1 - \frac{3\gamma}{4}} \mathbb{1}_2, \ E_2^{(1)} = \sqrt{\frac{\gamma}{4}} \sigma_x$$

$$E_3^{(1)} = \sqrt{\frac{\gamma}{4}} \sigma_y, \ E_4^{(1)} = \sqrt{\frac{\gamma}{4}} \sigma_z, \tag{A4}$$

where $\sigma_{x/y/z}$ are the Pauli matrices. Note that we adopted a scaling of the parameter γ for symmetric depolarising noise, such that at $\gamma = 1$ the output state is the completely mixed state $\frac{1}{2} \mathbb{1}_2$. In the Bloch sphere representation of a one-qubit state, pure states are represented as living on the surface of the sphere, while mixed states live in the interior, with the completely mixed state at the center. A symmetric depolarising channel can therefore be visualized as a uniform contraction of the Bloch vectors towards the center.

For a product channel, the set of Kraus operators is given by the tensor product of the set of single-qubit Kraus operators:

$$\{E_i^{(n)}\} = \{E_i^{(1)}\}^{\otimes N}. \tag{A5}$$

For example, the Kraus operators for a local dephasing channel on two qubits are:

$$E_1^{(2)} = E_1^{(1)} \otimes E_1^{(1)}, \ E_2^{(2)} = E_1^{(1)} \otimes E_2^{(1)},$$

$$E_3^{(2)} = E_2^{(1)} \otimes E_1^{(1)}, \ E_4^{(2)} = E_2^{(1)} \otimes E_2^{(1)}. \tag{A6}$$

To apply a product channel using the formula in Eq. (A1) would hence require the summation of exponentially many terms in the number of qubits. In practice, on the quantum simulator this is done by working in the Pauli picture. [58] Specifically, we can represent a general quantum state as a vector in the Pauli basis, and a general quantum channel as a matrix, also known as Pauli transfer matrix. Therefore, the application of a quantum channel, including noise channels, is reduced to a matrix multiplication between the corresponding Pauli transfer matrix and the vectorised quantum state. [59]

APPENDIX B: Linear model of noise propagation

In the interleaved noise model introduced in Sec. II B we consider a quantum circuit that we split in d layers. In the absence of noise, the unitary operation applied by the circuit can be written as:

$$U = U_d U_{d-1} \cdots U_2 U_1, \tag{B1}$$

where U_i is the ith layer unitary operator. In the noisy case we move to the density matrix picture for mixed states, and use maps between density operators (or superoperators) to represent operations on these mixed states, which are defined by $\mathcal{U}(\rho) = U\rho U^\dagger$. This represents a noiseless quantum channel. As outlined in the main text, this corresponding noisy quantum channel is given by

$$\tilde{\mathcal{U}} = \Lambda_d \mathcal{U}_d ... \Lambda_1 \mathcal{U}_1. \tag{B2}$$

1. Example: global depolarising noise

We first consider the case where all the noise channels are given by global symmetric depolarising noise on N qubits, since in this case we can write down an exact model for the noise propagation. The global symmetric depolarising channel is defined as acting on any state ρ_{in} as follows:

$$\Lambda(\rho_{\text{in}}) = (1-\gamma)\rho_{\text{in}} + 2^{-N}\gamma \mathbb{1}, \tag{B3}$$

where $\mathbb{1}$ is the identity operator on N qubits (in matrix form it is of dimension $2^N \times 2^N$). The effect of the channel is therefore to replace any state with the completely mixed state with probability γ.

If the circuit under consideration consists of d such channels interleaved with unitary operations, the probability of the state not being affected by any channel is $(1-\gamma)^d$. Any other outcome will lead to the final state

being maximally mixed, as the maximally mixed state is invariant under the action of any quantum channel. The final state of the circuit can therefore be written as

$$\rho = (1-\gamma)^d \mathcal{U}(\rho_{\text{in}}) + 2^{-N}(1-(1-\gamma)^d)\mathbb{1}, \quad \text{(B4)}$$

where $\mathcal{U}$ is the noiseless circuit unitary map.

We can quantify the difference between the state produced by the noisy circuit, ρ, and the state produced without noise, ρ_{id}, using the infidelity, defined in Eq. (9). From now on, the subscript id will universally refer to a quantity evaluated on an ideal, noiseless circuit. For the global symmetric depolarising noise operator in Eq. (B4) we obtain

$$
\begin{aligned}
&R(\rho_{\text{id}}, \rho)\\
&= 1 - \text{Tr}[\rho_{\text{id}}((1-\gamma)^d \rho_{\text{id}} + 2^{-N}(1-(1-\gamma)^d)\mathbb{1})]\\
&= 1 - (1-\gamma)^d \text{Tr}[\rho_{\text{id}}^2] - 2^{-N}(1-(1-\gamma)^d)\text{Tr}[\rho_{\text{id}}]\\
&= \left(1-(1-\gamma)^d\right)\left(1-2^{-N}\right), \quad \text{(B5)}
\end{aligned}
$$

where we use the fact that ρ_{id} is a pure state, and so $\text{Tr}[\rho_{\text{id}}^2] = 1$. Expanding at first order in γ we get

$$R(\rho_{\text{id}}, \rho) = (1-2^{-N})\, d\, \gamma + O(\gamma^2). \quad \text{(B6)}$$

This means that, in the low noise regime, the infidelity increases approximately linearly in both γ and the number of layers d, with a prefactor that is effectively 1 for large systems ($N \to \infty$).

Note that a global symmetric depolarising channel is not equivalent a product of single-qubit depolarising channels, and therefore the relation derived in Eq. (B6) does not apply for the product channel noise model that we employ in this work. However, the simple result of Eq. (B6) provides the motivation to seek an analogous equation for more general noise models.

2. Infidelity propagation for general noise channels

We start by expanding the action of a general N-qubit noise channel on a state ρ to first order as

$$\Lambda(\rho) = (1-\gamma)\rho + \gamma\lambda(\rho) + O(\gamma^2). \quad \text{(B7)}$$

Here λ is a map that represents the first-order action of the channel, which for global symmetric depolarising channels is simply $\lambda(\rho) = 2^{-N}\mathbb{1} \;\forall\rho$. Note that the matrix $\lambda(\rho)$ is, in general, not a valid quantum state. We term λ the *linear action* of the noise channel. In the expansion we take into account the fact that for a general noise channel there could be higher order effects in the local noise parameter γ. Since the gate noise level in current quantum computers has $\gamma \ll 1$ [35], we expect the linear action to always dominate, and hence that one can neglect higher order terms.

We introduce the partially noisy circuit:

$$\tilde{\mathcal{U}}_{:i:} := \mathcal{U}_d \cdots \mathcal{U}_{i+1}\lambda_i \mathcal{U}_i \cdots \mathcal{U}_1, \quad \text{(B8)}$$

obtained by inserting a single noise channel in the noiseless circuit and keeping only the terms at order γ. Applying the noisy circuit onto an initial state ρ_0 and expanding in $O(\gamma)$, it reduces to:

$$
\begin{aligned}
\rho &\approx \left[(1-\gamma d)\mathcal{U} + \gamma\sum_{i=1}^{d}\tilde{\mathcal{U}}_{:i:}\right](\rho_0)\\
&= \rho - \gamma\sum_{i=1}^{d}(\mathcal{U} - \tilde{\mathcal{U}}_{:i:})(\rho_0). \quad \text{(B9)}
\end{aligned}
$$

The resulting difference to the noiseless result is an additive contribution proportional to γ.

Now we combine the first-order expansion with the layered noise model, and apply it to the variational problem of maximising the fidelity to a pure target state ρ_{T}. In this case, the fidelity between the noisy state output by the circuit after optimisation ρ_{opt} and the target state simplifies to $\text{Tr}(\rho_{\text{T}}\rho_{\text{opt}})$. In general, the circuit may not reproduce the ideal target state ρ_{T} exactly. We therefore have

$$
\begin{aligned}
&F(\rho_{\text{T}}, \rho_{\text{opt}}) = \text{Tr}(\rho_{\text{T}}\rho_{\text{opt}})\\
&\approx \text{Tr}[\rho_{\text{T}}\rho_{\text{opt,id}} - \rho_{\text{T}}\gamma\sum_{i=1}^{d}(\mathcal{U} - \tilde{\mathcal{U}}_{:i:})(\rho_0)]\\
&= F(\rho_{\text{T}}, \rho_{\text{opt,id}}) - \gamma\sum_{i=1}^{d}(F(\rho_{\text{T}}, \rho_{\text{opt,id}}) - F(\rho_{\text{T}}, \tilde{\mathcal{U}}_{:i:}(\rho_0))).
\end{aligned}
$$
$$\text{(B10)}$$

We can now define the *relative infidelity* as

$$R_{\text{rel}}(\rho_{\text{T}}; \rho_{\text{opt,id}}, \rho_{\text{opt}}) := F(\rho_{\text{T}}, \rho_{\text{opt,id}}) - F(\rho_{\text{T}}, \rho_{\text{opt}}). \quad \text{(B11)}$$

We now define the state at the ith layer $\rho_{\text{opt},i}$, as well as the target state "back-propagated" to ith layer as

$$\rho_{\text{T},i} := \mathcal{U}_{:i}^{-1}\rho_{\text{T}}, \quad \text{(B12)}$$

where we use the fact that $\mathcal{U}_{:i}$ is a unitary channel. With this definition and the invariance of fidelity under unitary transformations, we can write Eq. (B10) as:

$$R_{\text{rel}}(\rho_{\text{T}}; \rho_{\text{opt,id}}, \rho_{\text{opt}}) \approx \gamma\sum_{i=1}^{d}R_{\text{rel}}(\rho_{\text{T},i}; \rho_{\text{opt},i}, \lambda(\rho_{\text{opt},i})). \quad \text{(B13)}$$

This result shows that the relative infidelity is approximately linear in the number of noise channels d and the local noise parameter γ. Each noise channel contributes a factor that depends on the target state ρ_{T}, the state through the channel $\rho_{\text{opt},i}$ and its image under the linear action $\lambda(\rho_{\text{opt},i})$. Note that this is consistent with the result for the symmetric depolarising channel in the previous section, by taking

$$R_{\text{rel}}(\rho_i; \rho_i, \lambda(\rho_i)) = R(\rho_i, \lambda(\rho_i)) = 1 - 2^{-n} \;\forall\rho_i. \quad \text{(B14)}$$

Importantly, our total infidelity is state dependent. Hence, the path through the space of states that the circuit takes is very important for the noise resilience of the algorithm. If two different paths lead to the same state in the noiseless case, they may still have very different infidelities in the noisy case, as dictated by these intermediate states.

To simplify the result and compare it to literature, one can estimate the average infidelity of a noise channel over all states, R_i, and with it obtain a very rough estimate for the total infidelity. First let us assume that the circuit can indeed reach the target state, such that we have

$$R_{\mathrm{rel}}(\rho_{,i}; \rho_i, \lambda(\rho_i)) = R(\rho_i, \lambda(\rho_i)) = 1 - \mathrm{Tr}[\rho_i \, \lambda(\rho_i)] \tag{B15}$$

and

$$R_{\mathrm{rel}}(\rho_{\mathrm{id}}; \rho_{\mathrm{id}}, \rho) = R(\rho_{\mathrm{id}}, \rho) = 1 - \mathrm{Tr}[\rho_{\mathrm{id}} \, \rho]. \tag{B16}$$

Now use the definition of the linear action in Eq. (B7) to write, to first order in γ,

$$\begin{aligned} \gamma R_{\mathrm{rel}}(\rho_i; \rho_i, \lambda(\rho_i)) &\approx \gamma - \mathrm{Tr}[\rho_i \, \Lambda(\rho_i)] + (1 - \gamma)\mathrm{Tr}[\rho_i^2] \\ &= 1 - \mathrm{Tr}[\rho_i \, \Lambda(\rho_i)]. \end{aligned} \tag{B17}$$

In accordance with literature, we define the gate infidelity as $R_i = 1 - \mathrm{Tr}[\rho_i \, \Lambda(\rho_i)]$. In this way with Eq. (B13) one obtains for the total infidelity of the circuit

$$R \approx \sum_{i=1}^{d} R_i. \tag{B18}$$

From this equation we obtain the total state-averaged circuit fidelity as function of the averaged fidelities of each noise channel, $F_i = 1 - R_i$, as

$$F = 1 - \sum_{i=1}^{d} R_i = \prod_{i=1}^{d} (1 - R_i) + O\left(R_i^2\right). \tag{B19}$$

Since the infidelities of a noise channels are small, we can neglect higher order terms, and obtain

$$F \approx \prod_{i=1}^{d} F_i, \tag{B20}$$

where we have used $F_i = 1 - R_i$. This last expression is widely used in literature. For example, in Ref. [29] the fidelity is assumed to behave in this way, and subsequent experimental results are shown to be consistent with this assumption. In Ref. [60] it is shown that the approximation error of a matrix product state representation of a quantum circuit is approximately multiplicative in the fidelity of a single layer. In Ref. [61], a similar reasoning shows that this multiplicative decay of fidelity applies for interleaved depolarising channels in QAOA circuits. The decay of the cost function with the number of noisy channel that follows from these equations was rigorously proven in Ref. [24] for local symmetric depolarising channels. Ref. [62] shows that infidelity grows

at most quadratically with the number of layers of a circuit, and for decoherent channels it is expected to grow linearly at first order, which is equivalent to what we have shown here. However, the measure used there is the gate infidelity over Haar-distributed states, which might not reflect the outcome of specific experiments [63].

3. Expected fidelity over an ensemble

Up to now we considered the infidelity for one target state. We extend this to the case where one is provided with an ensemble of n_{T} pure target states sampled from a distribution. As figure of merit we use the average optimal infidelity over the ensemble, which we defined in Eq. (11) as

$$\bar{R} = \langle R(\rho_{\mathrm{T}}, \rho_{\mathrm{opt}}) \rangle_{\mathrm{T}}, \tag{B21}$$

where ρ_{opt} is optimized with noise for each target state. An $\bar{R}$ of zero would imply that the quantum circuit can represent any N-qubit state exactly. In practice $\bar{R}$ is usually larger than zero even in the noiseless case due the inherent limitations of a given circuit, and the addition of noise further increases $\bar{R}$. We define the average optimal relative infidelity as

$$\bar{R}_{\mathrm{rel}} := \bar{R} - \bar{R}_{\mathrm{id}} = \langle R_{\mathrm{rel}}(\rho_{\mathrm{T}}; \rho_{\mathrm{opt,id}}, \rho_{\mathrm{opt}}) \rangle_{\mathrm{T}}, \tag{B22}$$

where as before the subscript id indicates the infidelity and density matrix evaluated without noise:

$$\bar{R}_{\mathrm{id}} = \langle R(\rho_{\mathrm{T}}, \rho_{\mathrm{opt,id}}) \rangle_{\mathrm{T}}. \tag{B23}$$

Next we derive an approximate relation for $\bar{R}_{\mathrm{rel}}$ allowing us to qualitatively understand its behavior and relate it to existing literature. Using Eq. (B13) and the fact that in our noise model all the noise channels are identical, we obtain

$$\bar{R}_{\mathrm{rel}} \approx \gamma \left\langle \sum_{i=1}^{d} R_{\mathrm{rel}}(\rho_{\mathrm{T},i}; \rho_{\mathrm{opt},i}, \lambda(\rho_{\mathrm{opt},i})) \right\rangle_{\mathrm{T}}. \tag{B24}$$

We now assume that at every level the noiseless $\rho_{\mathrm{T},i}$ and $\rho_{\mathrm{opt},i}$ are distributed identically to the final states ρ_{T} and ρ_{opt}. Thus we can write

$$\begin{aligned} \bar{R}_{\mathrm{rel}} &\approx \langle R_{\mathrm{rel}}(\rho_{\mathrm{T}}; \rho_{\mathrm{opt}}, \lambda(\rho_{\mathrm{opt}})) \rangle_{\mathrm{T}} \, \gamma \, d \\ &\approx \alpha \, \gamma \, d, \end{aligned} \tag{B25}$$

where we defined the constant of proportionality

$$\alpha := \langle R_{\mathrm{rel}}(\rho_{\mathrm{T}}; \rho_{\mathrm{opt}}, \lambda(\rho_{\mathrm{opt}})) \rangle_{\mathrm{T}}. \tag{B26}$$

Note that since $\lambda(\rho_{\mathrm{opt}})$ is generally not a valid quantum state by itself, since usually $\mathrm{Tr}[\lambda(\rho)] \neq 1$, α may also be greater than 1.

With the additional assumption about the distributions of the intermediate states, the variance of the relative infidelity can be estimated. Defining the variance of the optimal relative infidelity over the target states:

$$\Delta_{\rm rel}^2 := \langle R_{\rm rel}^2(\rho_{\rm T}; \rho_{\rm opt,id}, \rho_{\rm opt})\rangle_{\rm T} - \bar{R}_{\rm rel}^2$$
$$= \underset{\rm T}{\rm Var}\left(R_{\rm rel}(\rho_{\rm T}; \rho_{\rm opt,id}, \rho_{\rm opt})\right). \quad (B27)$$

Using Eq. (B13) and a standard property of the variance, within these approximations we obtain

$$\Delta_{\rm rel}^2 \approx \underset{\rm T}{\rm Var}\left(\gamma \sum_{i=1}^{d} R_{\rm rel}(\rho_{\rm T,i}; \rho_{\rm opt,i}, \lambda(\rho_{\rm opt,i}))\right)$$
$$= \gamma^2 \underset{\rm T}{\rm Var}\left(\sum_{i=1}^{d} R_{\rm rel}(\rho_{\rm T,i}; \rho_{\rm opt,i}, \lambda(\rho_{\rm opt,i}))\right). \quad (B28)$$

Now we must introduce some information about the correlation between states at every noise channel. Since states going through successive channels have a certain degree of similarity, in the sense that they are related by a short-depth sequence of unitaries, it should be expected that they retain a high level of correlation. Therefore, we assume that the states at every noise channel $\{\rho_{\rm opt,1}, \rho_{\rm opt,2}, \cdots, \rho_{\rm opt,d}\}$ are perfectly correlated. It follows that:

$$\Delta_{\rm rel}^2 \approx \beta\, \gamma^2 d^2, \quad (B29)$$

with

$$\beta := \underset{\rm T}{\rm Var}(R_{\rm rel}(\rho_{\rm T}; \rho_{\rm opt}, \lambda(\rho_{\rm opt}))). \quad (B30)$$

Note that, if instead we assumed that the states are uncorrelated with each other, we would get $\Delta_{\rm rel}^2 \approx \beta\,\gamma^2 d$.

Eqs. (B25) and (B29) define the stochastic model for noise propagation.

4. Estimation of α and β

In this section we show that α and β can be estimated with knowledge of only the distribution of target states and the noise channel properties. First of all, we note that both constants depend on the output state $\rho_{\rm opt}$, which depends on the capability of the circuit to approximate the target state. Since we wish to remove the dependence on the circuit entirely, we use as an approximation $\lambda(\rho_{\rm opt}) \approx \lambda(\rho_{\rm T})$. This is justified by the fact that, given a sufficiently expressive circuit, $\rho_{\rm opt}$ will not be much different from $\rho_{\rm T}$. Therefore, using the definitions in Eqs. (B26) and (B30) the parameters can be estimated from the target states as

$$\alpha \approx \langle R_{\rm rel}\left(\rho_{\rm T}; \rho_{\rm T}, \lambda(\rho_{\rm T})\right)\rangle_{\rm T} = \langle R\left(\rho_{\rm T}, \lambda(\rho_{\rm T})\right)\rangle_{\rm T}, \quad (B31)$$
$$\beta \approx \underset{\rm T}{\rm Var}(R(\rho_{\rm T}, \lambda(\rho_{\rm T}))), \quad (B32)$$

where we can switch the relative infidelity for regular infidelity as we have removed all dependence on the circuit.

Using Eq. (B7) and taking the derivative of $\Lambda(\rho)$ about $\gamma = 0$ we obtain

$$\left.\frac{d\Lambda(\rho)}{d\gamma}\right|_{\gamma=0} = \lambda(\rho) - \rho \;\; \rightarrow \;\; \lambda(\rho) = \left.\frac{d\Lambda(\rho)}{d\gamma}\right|_{\gamma=0} + \rho. \quad (B33)$$

Therefore we can write

$$R(\rho_{\rm T}, \lambda(\rho_{\rm T})) = 1 - {\rm Tr}\left(\rho_{\rm T}\lambda(\rho_{\rm T})\right)$$
$$= 1 - {\rm Tr}\left(\rho_{\rm T}\left(\left.\frac{d\Lambda(\rho_{\rm T})}{d\gamma}\right|_{\gamma=0} + \rho_{\rm T}\right)\right)$$
$$= 1 - {\rm Tr}\left(\rho_{\rm T}\left.\frac{d\Lambda(\rho_{\rm T})}{d\gamma}\right|_{\gamma=0}\right) - {\rm Tr}\left(\rho_{\rm T}^2\right)$$
$$= -{\rm Tr}\left(\rho_{\rm T}\left.\frac{d\Lambda(\rho_{\rm T})}{d\gamma}\right|_{\gamma=0}\right). \quad (B34)$$

Substituting back into Eqs. (B31) and (B32) we obtain

$$\alpha \approx -\left\langle\left.\frac{d}{d\gamma}{\rm Tr}(\rho_{\rm T}\Lambda(\rho_{\rm T}))\right|_{\gamma=0}\right\rangle_{\rm T}, \quad (B35)$$
$$\beta \approx \underset{\rm T}{\rm Var}\left(\left.\frac{d}{d\gamma}{\rm Tr}(\rho_{\rm T}\Lambda(\rho_{\rm T}))\right|_{\gamma=0}\right). \quad (B36)$$

If states can be efficiently sampled from the distribution, the constants α and β can be estimated given an exact density matrix simulator. Given a sampled target state, the derivative can be evaluated in practice using finite differences as

$$\left.\frac{d}{d\gamma}{\rm Tr}(\rho_{\rm T}\Lambda(\rho_{\rm T}))\right|_{\gamma=0} = \lim_{\epsilon \to 0}\frac{{\rm Tr}(\rho_{\rm T}\Lambda(\rho_{\rm T}))|_{\gamma=\epsilon} - 1}{\epsilon}. \quad (B37)$$

The results obtained with Eqs. (B35-B37) for the distribution of 4-qubits real states used in our simulations are shown in Table I.

Noise channel	α	β
Phase	0.888	0.00585
Amplitude	1.88	0.119
depolarising	2.78	0.0132

TABLE I: Estimated α and β for local noise channels on four qubits, averaged over the real Haar distribution states. 10000 randomly chosen states were used for the numerical estimation.

5. Scaling of α with the number of qubits

For an approximate scaling estimate of α with the number of qubits N, we consider simplified case of a target state being a tensor product of N independent identically distributed single-qubit states. With this as-

sumption we can write

$$
\begin{aligned}
\alpha &\approx -\frac{d}{d\gamma}\Big|_{\gamma=0} \mathrm{Tr}[\rho_{\mathrm{T}}\Lambda(\rho_{\mathrm{T}})] \\
&= -\frac{d}{d\gamma}\Big|_{\gamma=0} \mathrm{Tr}\left[\bigotimes_i \rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})\right] \\
&= -\frac{d}{d\gamma}\Big|_{\gamma=0} \prod_i \mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})] \\
&= -\sum_i \frac{d}{d\gamma}\Big|_{\gamma=0}\left(\mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})]\right)\prod_{j\neq i}\mathrm{Tr}[\rho_{\mathrm{T}}^{(j)}\Lambda^{(j)}|_{\gamma=0}(\rho_{\mathrm{T}}^{(j)})] \\
&= -\sum_i \frac{d}{d\gamma}\Big|_{\gamma=0}\mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})] \\
&= \sum_{i=1}^{N}\alpha^{(i)}, \tag{B38}
\end{aligned}
$$

where we defined the single-qubit quantity:

$$
\alpha^{(i)} := -\frac{d}{d\gamma}\Big|_{\gamma=0}\mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})]. \tag{B39}
$$

Since we assumed that the single-qubit product states are identically distributed, these terms are equal for all i and we obtain a scaling of $\alpha \sim O(N)$. Thus, we expect the effect of noise to grow linearly with the number of qubits.

APPENDIX C: Additional results

1. VQE for noise levels varying across qubits

In this subsection we show results for 2-qubit VQE simulations as done in Sec. III A, but for the case where the two qubits suffer from very unequal noise levels. We consider the case of one qubit having 10 times the noise of the other one. Analogously to the equal noise results in Fig. 2, we plot the converged energy, fidelity and concurrence with respect to the noise parameter γ, which here refers to the most noisy qubit. We do this for the three noise channels (phase, amplitude, depolarising), and focus only on the 4-parameter circuit (Fig. 1(c)). There

are two possible configurations: the first is where the most noisy qubit is the top qubit in Fig. 1(c), which we denote as circuit 0; the second is where the bottom qubit is more noisy, which we denote as circuit 1.

The results are presented in Fig. 11, and show that for phase and symmetric depolarising noise there is no difference between the two circuits, while for amplitude damping the difference is substantial. In particular, the local minima branch in very distinct directions in the two circuits. This indicates that the degree to which amplitude damping noise breaks the parameter degeneracies, leading to different local minima with increasing γ, is sensitive to the relative strength of the noise on the different qubits.
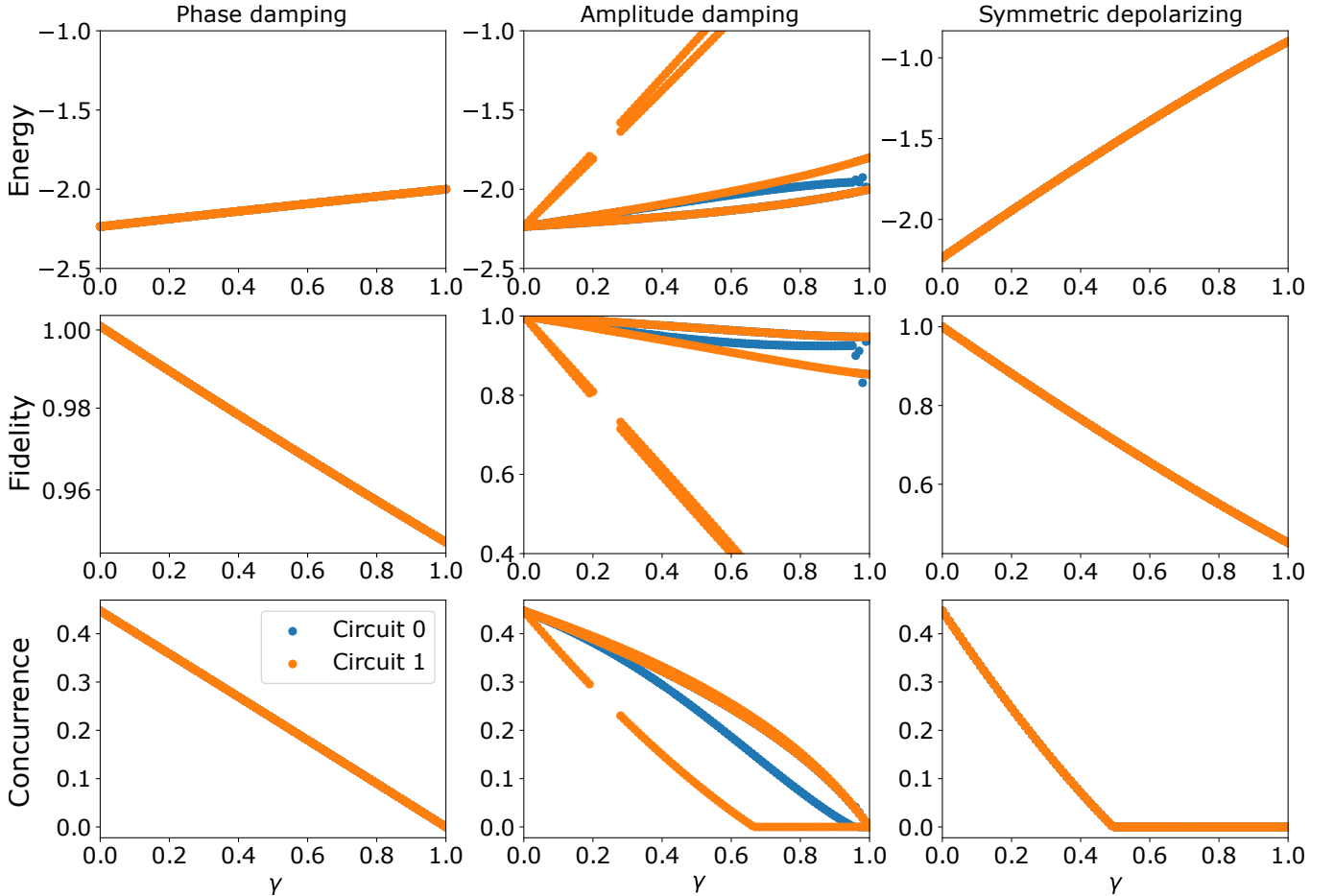

FIG. 11: Energy, fidelity and concurrence for the optimal energy minima found for the 2-qubit 4-parameter circuit in Fig. 1c, and for three different types of noise. In these simulations one qubit has a noise level 10 times less than the other: in circuit 0 the first qubit is more noisy, in circuit 1 the second qubit. The columns represent different types of noise channels, the rows the different measures of states quality.

2. Relative infidelity for different noise channels

We show the results for relative infidelity for amplitude damping and symmetric depolarising noise, for the target state optimization experiments. The results are similar to those for phase damping noise, presented in the main text (Fig. 10). There is a good agreement between the stochastic model and the numerical results for low noise parameter values. The model follows more closely the non-reoptimised results, which in the symmetric depolarising case match almost exactly the reoptimised results. Furthermore, the agreement improves with the number of layers, indicating that the assumptions of random intermediate states are more appropriate for deeper circuits.

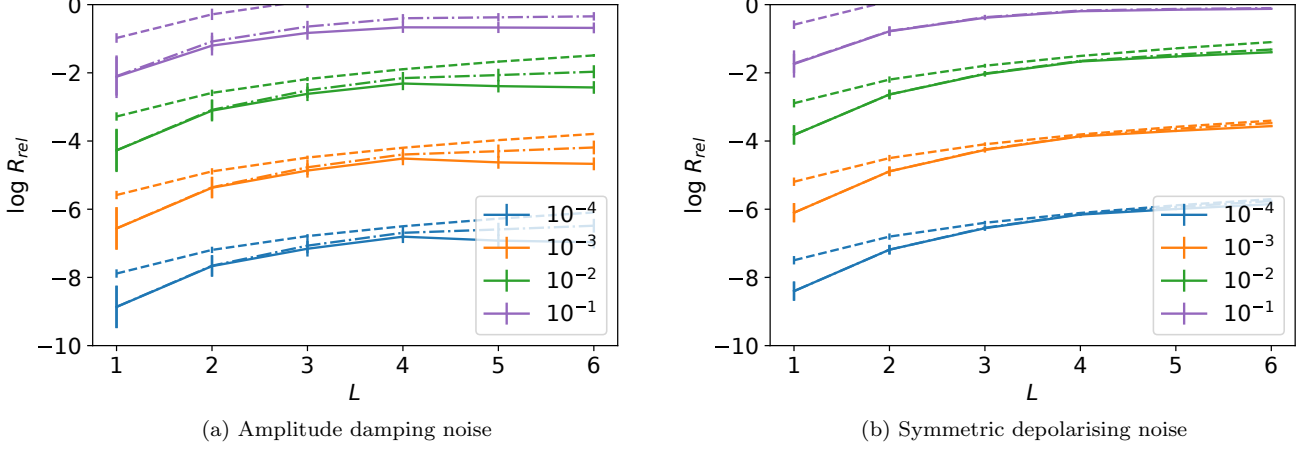(a) Amplitude damping noise

(b) Symmetric depolarising noise

FIG. 12: Relative infidelity vs layers, for a) amplitude damping noise and b) symmetric depolarising noise, for different values of the noise level γ. The solid curves are for noise-aware reoptimised parameters, the dash-dotted curves are for parameters optimised in the absence of noise (non-reoptimised), and the dashed curves indicate the results for the linear noise model. Each point shows the average over 1000 target states, and the verical bars at each point indicate the standard deviation.

3. Noise-induced state transitions for further random states

In Fig. 10 we show how fidelity and concurrence as function of γ abruptly change slope at at threshold γ value for one randomly chosen target state. To show that this behavior is general, in Fig. 13 we present analogous results for four more randomly chosen target states. We also consider different circuit depths. The transitions are overall very different to one another, at times discontinuous in their fidelity and at other times smooth and barely

detectable. In these latter cases, the concurrence acts as a clearer indicator for such transitions. The varied phenotype of transitions suggests that the phenomenon is complex and depends heavily on the circuit and the chosen target state. Indeed, in some situations (see the last row of Fig. 13) no sharp transitions are observed, and both measures of state quality vary smoothly. The results illustrate that generally one always observes critical noise level thresholds, and that these can be either abrupt and discontinuous or else smoothed out in a continuous way.
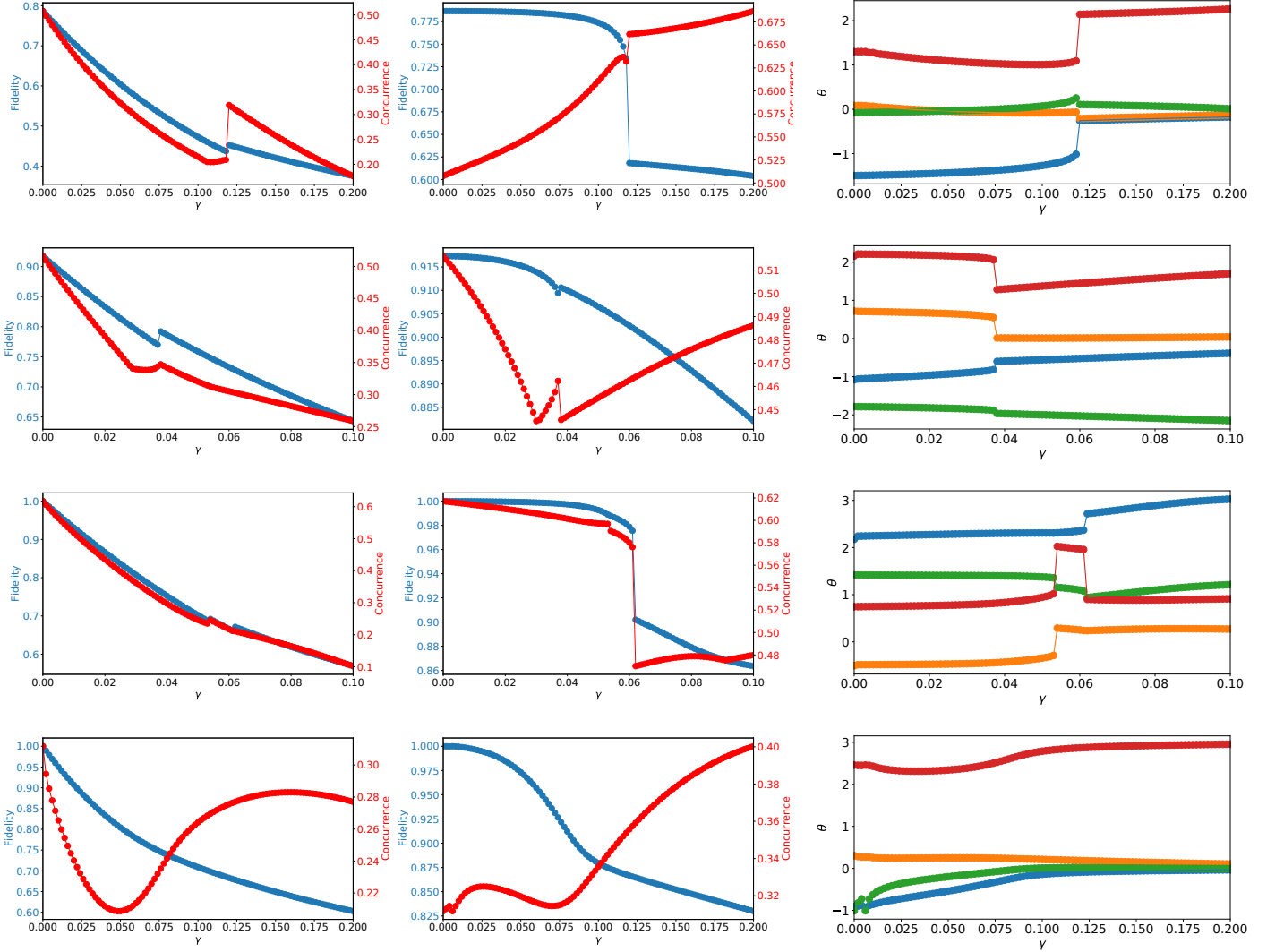

FIG. 13: Evidence of noise-induced state transitions for different random states. Various measures are plotted at convergence for a range of values of the phase noise parameter γ. The plotted data are described in more detail in fig. 10. Left panels: fidelity and concurrence at convergence, evaluated with noise; centre panels: fidelity and concurrence evaluated without noise, but for the converged optimized angles at every noise level; right panels: value of four of the circuit rotation parameters at convergence. The first two rows have a circuit of depth $L = 3$, the last two have $L = 4$.

[1] J. Preskill, Quantum computing in the nisq era and beyond, Quantum **2**, 79 (2018).

[2] A. Peruzzo, J. McClean, P. Shadbolt, M. H. Yung, X. Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5** (2014).

[3] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 10.1088/1367-2630/18/2/023023 (2016).

[4] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, Variational quantum algorithms for discovering Hamiltonian spectra, Phys. Rev. A **99**, 62304 (2019).

[5] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, Quantum **3**, 140 (2019).

[6] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Variational Quantum Fidelity Estimation, Quantum **4**, 248 (2020).

[7] M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles, Variational quantum state eigensolver, arXiv:2004.01372 .

[8] C. Cirstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, NPJ Quantum Inf. **6**, 1 (2020).

[9] X. Xu, J. Sun, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational algorithms for linear algebra, arXiv:1909.03898 .

[10] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices, Phys. Rev. X **10**, 021067 (2020).

[11] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, Quantum **3**, 191 (2019).

[12] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, *et al.*, Quantum optimization using variational algorithms on near-term quantum devices, Quantum Science and Technology **3**, 030503 (2018).

[13] K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, Noise resilience of variational quantum compiling, New J. Phys. **22**, 043006 (2020).

[14] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm, Phys. Rev. X **8**, 11021 (2018).

[15] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, Phys. Rev. A **95**, 1 (2017).

[16] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Hybrid Quantum-Classical Approach to Correlated Materials, Phys. Rev. X **6**, 31045 (2016).

[17] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, Phys. Rev. Lett. **118**, 150503 (2017).

[18] P. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, *et al.*, Scalable Quantum Simulation of Molecular Energies, Phys. Rev. X **6**, 31007 (2016).

[19] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[20] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[21] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, Phys. Rev. Lett. **122**, 180501 (2019).

[22] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, Phys. Rev. X **7**, 21050 (2017).

[23] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 31027 (2018).

[24] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-Induced Barren Plateaus in Variational Quantum Algorithms, arXiv:2007.14384 .

[25] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[26] B. Schumacher, Sending entanglement through noisy quantum channels, Phys. Rev. A **54**, 2614 (1996).

[27] L. Zhang, Y. Yu, C. Zhu, and C. Pei, Noise tailoring for quantum circuits via unitary 2t-design, Sci. Rep. **9**, 1 (2019).

[28] A. Bassi and D. A. Deckert, Noise gates for decoherent quantum circuits, Phys. Rev. A **77**, 10.1103/PhysRevA.77.032323 (2008).

[29] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[30] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2010).

[31] X. Hu, R. de Sousa, and S. D. Sarma, Decoherence and dephasing in spin-based solid state quantum computers, in *Foundations Of Quantum Mechanics In The Light Of New Technology: ISQM—Tokyo'01* (World Scientific, 2002) pp. 3–11.

[32] N. P. Sawaya, M. Smelyanskiy, J. R. McClean, and A. Aspuru-Guzik, Error sensitivity to environmental noise in quantum circuits for chemical state preparation, J. Chem. Theory Comput. **12**, 3097 (2016).

[33] B. Rost, B. Jones, M. Vyushkova, A. Ali, C. Cullip, A. Vyushkov, and J. Nabrzyski, Noisy Simulation of Quantum Beats in Radical Pairs on a Quantum Computer, arXiv:2001.00794.

[34] S. S. Tannu and M. K. Qureshi, A Case for Variability-Aware Policies for NISQ-Era Quantum Computers, arXiv:1805.10224.

[35] M. N. Lilly and T. S. Humble, Modeling Noisy Quantum Circuits Using Experimental Characterization, arXiv:2001.08653.

[36] R. Jozsa, Fidelity for Mixed Quantum States, J. Mod. Opt. **41**, 2315 (1994).

[37] S. Hill and W. K. Wootters, Entanglement of a pair of quantum bits, Phys. Rev. Lett. **78**, 5022 (1997).

[38] V. Coffman, J. Kundu, and W. K. Wootters, Distributed entanglement, Phys. Rev. A **61**, 052306 (2000).

[39] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, Phys. Rev. Lett. **80**, 2245 (1998).

[40] H. Lyyra, G. Karpat, C. F. Li, G. C. Guo, J. Piilo, and S. Maniscalco, Symmetry in the open-system dynamics of quantum correlations, Sci. Rep. **7**, doi.org/10.1038/s41598-017-08457-1 (2017).

[41] W. K. Wootters, Entanglement of formation and concurrence, Quantum Inf. Comput. **1**, 27 (2001).

[42] P. J. Love, A. M. Van Den Brink, A. Y. Smirnov, M. H. Amin, M. Grajcar, E. Il'ichev, A. Izmalkov, and A. M. Zagoskin, A characterization of global entanglement, Quantum Inf. Process. **6**, 187 (2007).

[43] M. Yönaç, T. Yu, and J. H. Eberly, Pairwise concurrence dynamics: a four-qubit model, J. Phys. B **40**, S45 (2007).

[44] A. Meill and D. A. Meyer, Pairwise concurrence in cyclically symmetric quantum states, Phys. Rev. A **100**, 10.1103/PhysRevA.100.042318 (2019).

[45] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm, NPJ Quantum Inf. **6** (2020).

[46] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[47] M. Heyl, A. Polkovnikov, and S. Kehrein, Dynamical quantum phase transitions in the transverse-field Ising model, Phys. Rev. Lett. **110**, 135704 (2013).

[48] I. Rungger, N. Fitzpatrick, H. Chen, C. H. Alderete, H. Apel, A. Cowtan, A. Patterson, D. M. Ramo, Y. Zhu, N. H. Nguyen, E. Grant, S. Chretien, L. Wossnig, N. M. Linke, and R. Duncan, Dynamical mean field theory algorithm and experiment on quantum computers, arXiv:1910.04735 .

[49] C. G. Broyden, The Convergence of a Class of Double-rank Minimization Algorithms: 2. The New Algorithm, IMA J. Appl. Math. **6**, 222 (1970).

[50] R. Fletcher, A new approach to variable metric algorithms, Comput. J. **13**, 317 (1970).

[51] D. Goldfarb, A Family of Variable-Metric Methods Derived by Variational Means, Math. Comput. **24**, 23 (1970).

[52] D. F. Shanno, Conditioning of Quasi-Newton Methods for Function Minimization, Math. Comput. **24**, 647 (1970).

[53] E. Fontana, M. Cerezo, A. Arrasmith, I. Rungger, and P. J. Coles, Optimizing parametrized quantum circuits via noise-induced breaking of symmetries, arXiv:2011.08763 (2020).

[54] W. Research, CircularRealMatrixDistribution - Wolfram Language Documentation, https://reference.wolfram.com/language/ref/CircularRealMatrixDistribution.html.

[55] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms, Adv. Quantum Technol. **2**, 1900070 (2019).

[56] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, Quantum Sci. Technol. **4**, 1 (2019).

[57] B. Tarasinski, Quantumsim, https://github.com/quantumsim/quantumsim/tree/stable/v0.2 (2016).

[58] B. Tarasinski, Pauli bases - quantumsim documentation, https://quantumsim.gitlab.io/architecture/pauli.html (2018).

[59] D. Greenbaum, Introduction to quantum gate set tomography, arXiv:1509.02921 (2015).

[60] Y. Zhou, E. M. Stoudenmire, and X. Waintal, What limits the simulation of quantum computers?, (2020), arXiv:2002.07730.

[61] J. Marshall, F. Wudarski, S. Hadfield, and T. Hogg, Characterizing local noise in QAOA circuits, IOP SciNotes **1**, 025208 (2020).

[62] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, Bounding the average gate fidelity of composite channels using the unitarity, New J. Phys. **21**, 10.1088/1367-2630/ab1800 (2019).

[63] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, Bounding quantum gate error rate based on reported average fidelity, New J. Phys. **18**, 12002 (2015).