Ability of error correlations to improve the performance of variational quantum algorithms

Joris Kattemölle and Guido Burkard
Department of Physics, University of Konstanz, D-78457 Konstanz, Germany

The quantum approximate optimization algorithm (QAOA) has the potential of providing a useful quantum advantage on noisy intermediate-scale quantum (NISQ) devices. The effects of uncorrelated noise on variational quantum algorithms such as QAOA have been studied intensively. Recent experimental results, however, show that the errors impacting NISQ devices are significantly correlated. We introduce a model for both spatially and temporally (non-Markovian) correlated errors based on classical environmental fluctuators. The model allows for the independent variation of the marginalized spacetime-local error probability and the correlation strength. Using this model, we study the effects of correlated stochastic noise on QAOA. We find evidence that the performance of QAOA improves as the correlation time or correlation length of the noise is increased at fixed local error probabilities. This shows that noise correlations in themselves need not be detrimental for NISQ algorithms such as QAOA.

I. INTRODUCTION

Quantum computers hold the promise of outperforming classical computers on tasks such as the simulation of quantum-mechanical systems [1,2], factoring [3], unstructured database search [4], and solving linear systems of equations [5]. However, it is much harder to shield qubits, the fundamental building blocks of quantum computers, from environmental noise than it is to shield classical bits [6], leading to errors in quantum computations. These errors can be detected and corrected, provided that error probabilities remain below a certain threshold, as is proven by threshold theorems for fault-tolerant quantum computation [7–10]. Fault-tolerant quantum computation remains possible in the presence of some forms of spatially and temporally correlated errors [10–14] but breaks down in others [15]. In any case, the quantum overhead required for fault-tolerant quantum computation is currently prohibitively large.

Nevertheless, current pre-error-corrected noisy intermediate-scale quantum (NISQ) [16] computers can already outperform classical computers on some tasks and have hence shown a quantum advantage [17,18]. However, the tasks for which a quantum advantage has been demonstrated are artificial and have no known applications. Thus, the next milestone in the field will be the demonstration of a *useful* quantum advantage.

Hybrid quantum-classical variational quantum algorithms (VQAs) have the potential of showing a useful quantum advantage on NISQ devices [19]. In these algorithms, a parametrized quantum state, known as the ansatz, is prepared using a parametrized quantum circuit. Subsequently, the expectation value of some observable (which depends on the specific VQA) is estimated by repeated preparation and measurement of the ansatz. The parameters and the resulting expectation value are fed into a classical optimization algorithm, which in turn suggests new parameters. By repeating the above steps, the optimizer heuristically seeks parameters

that optimize the expectation value of the observable. The VQA returns the optimal value of the observable and the parameters at which it was attained. Using these parameters, the optimal ansatz state can be reconstructed, which allows for the extraction of additional classical data from the optimal ansatz state.

The quantum approximate optimization algorithm (QAOA) [20] is a VQA designed to (approximately) solve instances from a large set of optimization problems [21], including problems of practical relevance, such as portfolio optimization [22] and correlation clustering [23]. The cost function of the optimization problem is first mapped to an Ising-type Hamiltonian whose ground state corresponds
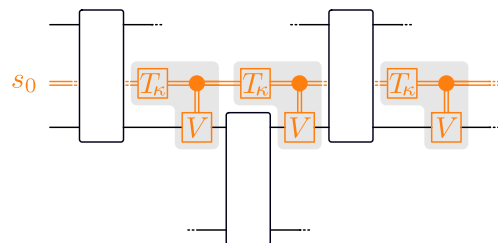

FIG. 1. The temporal fluctuator model. Each qubit interacts with an independent classical two-level fluctuator in the bath (orange, only one fluctuator shown). The fluctuator is initially excited with probability p, as described by the classical ensemble s_0. The operation T_κ resets the fluctuator to s_0 with probability $1 - \kappa$. We refer to κ as the correlation strength. The unitary error operator V is applied to the qubit if and only if the fluctuator is excited, leading to an error model with independent time-local error probability p and correlation time $0 \leqslant \tau = -1/\ln \kappa < \infty$ (in units of gate time). For a circuit of depth m, acting on n qubits, the full temporal fluctuator model is obtained by repeating the shaded area m times and the entire structure n times. Empty rectangles represent generic gates. At the end of the circuit, the fluctuator is discarded.

to the optimal solution of the optimization problem. This Hamiltonian is the observable being measured during the optimization loop. In the context of QAOA, the expectation value of the Hamiltonian, as a function of the variational parameters, is called the cost function landscape. QAOA seeks approximate ground states of the Hamiltonian by optimizing the cost function landscape and thereby finds approximate solutions to the optimization problem.

A common figure of merit for QAOA is the approximation ratio (AR), defined as the ratio of the cost function landscape at the best parameters found by QAOA to the global optimal value of the classical cost function. Along the lines of Ref. [24], to evaluate the effects of noise, a *noise-unaware* AR can be defined. The noise-unaware AR is obtained by first running QAOA in classical simulations with all noise turned off. The resulting optimal parameters are subsequently used to compute the expectation value of the Hamiltonian while the noise is turned on. An AR that is larger than the noise-unaware AR indicates that variational parameters can adapt to the presence of noise. We call this ability the *noise adaptivity* of the optimal parameters.

The behavior of VQAs such as QAOA under *local* errors has been studied extensively [24–32], showing that VQAs possess forms of resilience against coherent [27,28] and incoherent [24,29–32] errors. VQAs [29], specifically QAOA [33], are therefore thought to be ideal candidates for obtaining a useful quantum advantage.

Recent experimental results have highlighted the ubiquity of both spatially and temporally correlated errors in NISQ devices. Sources of spatially correlated errors include crosstalk [34,35], fluctuations of external, quasihomogeneous magnetic fields [36], and (cosmic) radiation [37,38]. Sources of temporally correlated errors include $1/f$ noise (where f denotes frequency) [39–42], reflections in drive lines [43], long-lived quasiparticle excitations [37], nuclear spins [44], and microscopic two-level systems [45–47]. In particular, Refs. [46,47] studied the stability of superconducting (transmon) qubits. Telegraph-like switching of relaxation times [47] and qubit frequency [46] were observed. Both works attribute these effects to single two-level systems local to each qubit.

Given the importance of VQAs, their resilience against local errors and the prevalence of correlated errors, an understanding of the effects of correlated errors on VQAs is needed. We make a first step in this direction by studying the effects of temporally correlated (non-Markovian) and spatially correlated errors on QAOA.

We introduce a toy model for temporally correlated errors (Fig. 1) that is inspired by the two-level systems local to each qubit of Refs. [46,47]. Our model captures the essence of correlated errors, which we consider to be the existence of a nontrivial error probability and correlation strength. To this end, we assign a classical two-level fluctuator to each qubit that causes a unitary error operation V on the associated qubit conditioned on the state of the fluctuator. The error probability p and correlation time τ of these errors, defined in Fig. 1, are determined by the internal evolution of the fluctuators. A significant advantage of this model is its ability to fully and independently control the marginalized, time-local error probability p and correlation time τ. Interchanging the roles of space and time, a model for spatially correlated errors is
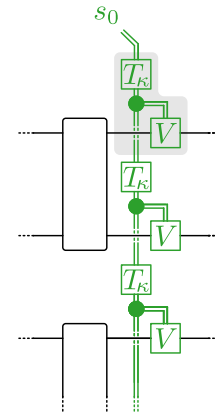

FIG. 2. The spatial fluctuator model. Definitions are as in Fig. 1. As opposed to Fig. 1, the fluctuator is displayed in green and travels through space rather than time. The error model for one circuit layer is generated by repeating the shaded area in the spatial direction n times, with n being the number of qubits. The full error model is obtained by repeating the obtained structure after each gate time. In the spatial fluctuator model, the correlation strength κ leads to an error correlation length $0 \leqslant \lambda = -1/\ln \kappa < \infty$ in units of interqubit distance.

obtained with local error probability p and correlation length λ (Fig. 2). Both τ and λ arise from the same fluctuator parameter κ that we call the *correlation strength*. Thus, we are able to treat temporally and spatially correlated errors on an equal footing. Limitations of our model are discussed in Sec. V of this paper.

We classically simulate the performance of QAOA under the fluctuator error models of Figs. 1 and 2 for various problem instances, error probabilities p, and correlation strengths κ. Our main results are the following.

Our first result is the observation of an *increase* of the performance of QAOA (as measured by the AR) as a function of the correlation strength κ at fixed spacetime-local marginalized error probabilities p across all cases studied. We analytically derive the effect of our fluctuator models on the AR to first order in the spacetime-local error probability p. Using this linear order theory, we show the increase in AR with correlation strength is explained by a counting argument, stating that a correlated error can happen in fewer ways than an uncorrelated error. Although a single, strongly correlated error can have a stronger effect on the output state than a single error happening at a single spacetime location, in all cases we studied this does not outweigh the fact that there are simply fewer ways in which a correlated error may happen.

A second, distinct result is the observation of a linear degradation of the performance of QAOA as a function of the error probability p around $p = 0$, noise adaptivity of QAOA, and the existence of critical error probabilities for this noise adaptivity, all for uncorrelated and correlated errors that do not break the symmetries of the cost function landscape. As p is increased away from $p = 0$, the AR and the noise-unaware AR decrease linearly and are essentially equal up until some critical p that depends on the fluctuator model (temporal or spatial), the correlation strength, and problem instance. After this critical p, the AR becomes better than the noise-unaware

AR. This divergence between the AR and noise-unaware AR coincides with an abrupt jump in optimal parameters that are otherwise essentially constant. This indicates that there are two competing, separated points in the cost function landscape and that one of them overtakes the other at the critical p.

Recently, similar linear decay, noise adaptivity of optimal parameters, and critical error probabilities were observed for a VQA by Fontana *et al.* [24]. Our results regarding these points go beyond the results of Fontana *et al.*, first because we show them for QAOA, and second because our results also hold for correlated errors. Finally, Fontana *et al.* attributed the noise adaptivity to the fact that their noise model breaks the symmetries of the cost function landscape. Our fluctuator error models do not break the symmetries of the cost function landscape, showing the existence of additional mechanisms for noise adaptivity.

The remainder of this paper is organized as follows: We give a more detailed introduction to QAOA in Sec. II. In Sec. III, we derive physical properties of our error model, such as correlation functions and marginalized error probabilities. In the same section, we derive the linear order effects of our fluctuator models on the cost function landscape. Subsequently, in Sec. IV, we describe our numerical methods and present and analyze our results, followed by a discussion in Sec. V.

II. QUANTUM APPROXIMATE OPTIMIZATION ALGORITHM

A wide class of optimization problems can be formulated using a quadratic cost function, $C(z) = \sum_{i<j}^{n} \omega_{ij} z_i z_j + \sum_{i=1}^{n} \omega_i z_i$, with $z_i \in \{-1, 1\}$ [21]. The goal is to find a $z = (z_1, \ldots, z_n)$ that optimizes C globally. Depending on the specific application, the optimization is a maximization or minimization of C. The cost function can readily be mapped to a Hamiltonian on n qubits,

$$H = \sum_{i<j}^{n} \omega_{ij} Z_i Z_j + \sum_{i=1}^{n} \omega_i Z_i, \tag{1}$$

with Z_i the Pauli-Z operator acting on qubit i. The mapping is such that $C(z) = \langle z|H|z\rangle$. The *Sherrington–Kirkpatrick* (SK) model encompasses all cost functions with $\omega_{ij} \in \{1, -1\}$ and $\omega_i = 0$. In this case, both C and H can be identified with the same undirected weighted graph with n nodes and adjacency matrix ω_{ij}. We do not assume the SK model in this section unless stated otherwise.

QAOA [20] bounds the optimal value of C by optimizing the cost function landscape $\tilde{C} : \mathbb{R}^{2r} \to \mathbb{R}$, with

$$\tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \langle \boldsymbol{\gamma}, \boldsymbol{\beta}|H|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle, \tag{2}$$

by a classical heuristic optimization method of choice. Here, $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$ is the ansatz quantum state, depending on $2r$ parameters. It is prepared on a quantum computer by

$$|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle = \prod_{k=r}^{1} S(\gamma_k, \beta_k)|+\rangle^{\otimes n}, \tag{3}$$

with $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, and the cycle

$$S(\gamma_k, \beta_k) = \exp\left(-i\frac{\beta_k}{2} \sum_{i=1}^{n} X_i\right) \exp\left(-i\frac{\gamma_k}{2} H\right)$$

$$= RX^{\otimes n}(\beta_k) \left[\prod_i RZ_i(\omega_i \gamma_k)\right]$$

$$\times \left[\prod_{i<j} RZZ_{ij}(\omega_{ij} \gamma_k)\right]. \tag{4}$$

Here, $RX(\alpha) = e^{-i\alpha X/2}$, $RZ_i(\alpha) = e^{-i\alpha Z_i/2}$, and $RZZ_{ij}(\alpha) = e^{-i\alpha Z_i Z_j/2}$. In this work, the order of the products is from left to right, so that in Eq. (3), $S(\gamma_1, \beta_1)$ is the first cycle to act on $|+\rangle^{\otimes n}$. To obtain an estimate for $\tilde{C}$, the ansatz state is prepared repeatedly, each time measuring H at the end of the circuit. Since H is diagonal in the computational basis, this measurement can be performed by measurements in the computational basis and classical postprocessing of the measurement outcomes.

On NISQ devices, the ansatz state will be described by a mixed state $\rho(\boldsymbol{\gamma}, \boldsymbol{\beta})$. In this work, the focus is on correlated errors during the ansatz circuit and we will therefore assume perfect estimation of $\tilde{C}$ throughout. That is, we use

$$\tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \text{tr}[\rho(\boldsymbol{\gamma}, \boldsymbol{\beta})H] \tag{5}$$

to compute the cost function landscape in our classical simulation of QAOA (Sec. IV).

For SK problem Hamiltonians ($\omega_{ij} \in \{1, -1\}$, $\omega_i = 0$), each cycle of the ansatz circuit requires one two-qubit gate RZZ between each pair of qubits. To accomplish this on quantum hardware without all-to-all connectivity, SWAP gates need to be inserted. On hardware with square-grid or line connectivity, the optimal way of doing so is by a SWAP network [48,49]. In the SWAP-network implementation of the ansatz circuit (Fig. 3), the order of the product of RZZ gates forms a brickwork structure, and a SWAP gate is inserted after each RZZ gate [50]. From $\omega_{ij} \in \{1, -1\}$ and the antiperiodicity of RZZ (and RZZ') with period 2π, it follows that the circuit cycle $S(\gamma_k, \beta_k)$ is invariant (up to an irrelevant overall sign) under $\gamma_k \mapsto \gamma_k \pm 2\pi$ for any k. As a consequence, the noiseless cost function landscape is fully periodic in γ_k with period 2π for any k. A similar statement holds for RX. Because of the properties of the circuit ansatz and the SK Hamiltonian, much stronger and nontrivial symmetry relations hold, which are derived in Appendix B. In the same appendix, we prove that these symmetries are not broken by spatiotemporally correlated Pauli error channels, which includes our fluctuator models if the unitary error operator V (defined in Figs. 1 and 2) is a Pauli operator.

Returning to the general case [Eq. (1)], a widely used figure of merit of optimization algorithms is the *approximation ratio* (AR),

$$AR = \frac{\tilde{C}(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)}{C^*}, \tag{6}$$

with $\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*$ being the optimal parameters as returned by the optimization algorithm, and C^* the true, global optimum of the cost function C, which is equal to the ground-state energy
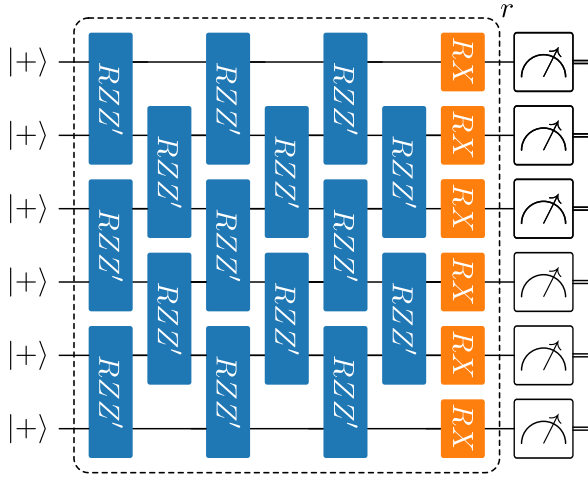
FIG. 3. SWAP-network implementation of the QAOA ansatz on $n = 6$ qubits in the absence of 1-local terms in the problem Hamiltonian ($\omega_i = 0$), with $RZZ' = RZZ \cdot$ SWAP and dependence on variational parameters suppressed. The ansatz cycle, marked by the dashed box, is applied to the initial state $|+\rangle^{\otimes n}$ a total of r times, each time with new parameters. The cost function landscape is obtained through repeated preparation, measurement, and classical postprocessing.

of H by construction. The algorithm has found the global optimum of any C if and only if $AR = 1$. If, in contrast, the ansatz state is maximally mixed, $\rho(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \mathbb{1}/2^n$, which amounts to randomly guessing measurement outcome bit stings, $AR = 0$ by the definition of $\tilde{C}$ [Eq. (5)] and the tracelessness of H [Eq. (1)]. In a noisy implementation of QAOA, the AR that is reached is in principle *noise aware*; noise may cause the optimum of $\tilde{C}$ to change and the optimal parameters may adapt to this change by classical heuristic optimization.

To assess the resilience of QAOA against correlated errors, in this work we also consider the *noise-unaware* AR. In classical simulation, where error probabilities can be changed at will, QAOA can be run without noise, obtaining the optimal parameters $\boldsymbol{\gamma}_0^*$, $\boldsymbol{\beta}_0^*$. These parameters are thus unaware of the effects of noise on the cost function landscape. Putting them into the noisy circuit, we obtain $\tilde{C}(\boldsymbol{\gamma}_0^*, \boldsymbol{\beta}_0^*)$. Thus, we define the *noise-unaware AR* as

$$AR_0 = \frac{\tilde{C}(\boldsymbol{\gamma}_0^*, \boldsymbol{\beta}_0^*)}{C^*}. \tag{7}$$

Using the noise-unaware AR, we define the *noise adaptivity* as

$$\Delta AR = AR - AR_0. \tag{8}$$

Given that the optimum found in the noiseless case and the optimum found in the noisy case are global, we have $\Delta AR \geqslant 0$, where $\Delta AR \neq 0$ indicates that the optimal parameters have adapted to the noise.

A quantity similar to AR_0 was introduced in Ref. [24]. There, instead of the expectation value of the Hamiltonian, the fidelity between the ansatz state and a target state was used as the cost function. Furthermore, the optimal noise-aware parameters were obtained by using the noise-unaware parameters as the initial point of optimization. This led the authors of Ref. [24] to call their noise-aware parameters the

reoptimized parameters. The advantage of reoptimization is that $\Delta AR \geqslant 0$ is guaranteed. In the current work, we find the optimal parameters for AR and AR_0 independently to prevent any bias of the noise-aware optimal parameters to points close to the noise-unaware parameters. Furthermore, our data (Sec. IV) obey $\Delta AR \geqslant 0$ without enforcing this property by reoptimization.

III. ERROR MODEL

The internal time evolution of a single fluctuator can be described by a two-state, discrete-time, and time-homogeneous Markov chain. (See, e.g., Refs. [51,52] for background information on Markov chains.) Using this formulation, we obtain the correlation time, time-local marginalized error probabilities, and the expected number of errors of the error models used in this paper (Figs. 1 and 2). Analytical first-order effects of *uncorrelated* errors in the context of variational quantum algorithms were studied before in Refs. [24–26]. We extend these methods to derive the analytical first-order effects of our correlated error models on QAOA.

In this section, we use terminology from the model of temporally correlated errors. Nonetheless, the results hold for the spatial fluctuator model if the state of the fluctuator after circuit layer i is reinterpreted as the state of the fluctuator after qubit i. In this case, the correlation time τ becomes the correlation length λ.

A. Two-level fluctuator

Consider a single classical two-level fluctuator with ground state $(1, 0)^{\mathrm{T}}$ and excited state $(0, 1)^{\mathrm{T}}$, with T being the transpose. If the fluctuator is excited with probability $0 \leqslant p \leqslant 1$, its state is described by the ensemble

$$s_0 = \begin{pmatrix} 1 - p \\ p \end{pmatrix}. \tag{9}$$

This does not represent a quantum state, but a classical ensemble equivalent to the classical mixed quantum state $\mathrm{diag}(s_0)$. We take s_0 as the initial state of the fluctuator. Define the random variable

$$B[(1, 0)^{\mathrm{T}}] = 0, \quad B[(0, 1)^{\mathrm{T}}] = 1, \tag{10}$$

which can, e.g., be imagined as the strength of a magnetic field caused by the fluctuator at its associated qubit.

At time t (in units of gate time), the state of the fluctuator, s_t, is retained with probability $0 \leqslant \kappa \leqslant 1$, and it is reset to s_0 with probability $1 - \kappa$. We refer to κ as the correlation strength. After this process, the state is given by $s_{t+1} = T s_t$, with T the transition matrix,

$$T = \begin{bmatrix} \kappa + (1 - \kappa)(1 - p) & (1 - \kappa)(1 - p) \\ (1 - \kappa)p & \kappa + (1 - \kappa)p \end{bmatrix}, \tag{11}$$

denoted by T_κ in Figs. 1 and 2.

To see that T describes the desired process, assume, for example, that at time t, the fluctuator is excited; $s_t = (0, 1)^{\mathrm{T}}$. After it is reset with probability $1 - \kappa$, there are two ways in which it can remain excited: either the fluctuator was not reset, which happens with probability κ, or the fluctuator was reset to s_0, but is excited merely because in the state s_0, the

fluctuator is excited with probability p. Thus, the overall probability that the fluctuator remains excited is $(0, 1)T(0, 1)^{\mathrm{T}} = T_{11} = \kappa + (1 - \kappa)p$. Other entries of T follow similarly. Note T_{ba} describes the probability to transition from the state a to b (in terms of the values that the random variable B can take).

The initial state s_0 is a steady state of T because T sends s_0 to s_0 both when the state is kept and when it is reset. (Explicit diagonalization is a simple way to check that this is the only nontrivial steady state). Thus, from an ensemble viewpoint, the Markov chain is trivial; $s_t = T^t s_0 = s_0$ for all t. Therefore, the marginalized, time-local probability to be in the excited state is p, independent of t and κ. Under the definition $B_t = B(s_t)$, it follows that $\mathbb{E}(B_t) = p$ for all t. Defining the random variable $B_{\mathrm{tot}} = \sum_{t'=0}^{t} B_{t'}$, it follows by the linearity of expectation values, which also holds for correlated random variables, that the expected number of times the fluctuator has been excited from $t = 0$ up to and including $t = t'$ is simply $(t + 1)p$, irrespective of κ.

The state of the fluctuator at time t is given by $s_t = T^t s_0$, where the tth matrix power of T is given by

$$T^t = \begin{pmatrix} 1 - p + p\kappa^t & 1 - p - \kappa^t + p\kappa^t \\ p - p\kappa^t & p + \kappa^t - p\kappa^t \end{pmatrix}. \quad (12)$$

Note that Eq. (11) is retrieved at $t = 1$. With a concise expression for T^t at hand, the correlator $\mathcal{C}(\Delta t) = \mathbb{E}(B_t B_{t+\Delta t}) - \mathbb{E}(B_t)\mathbb{E}(B_{t+\Delta t})$, which does not depend on t by time homogeneity of the Markov chain, is computed straightforwardly, yielding

$$\mathcal{C}(\Delta t) = 4p(1 - p)\kappa^{|\Delta t|}. \quad (13)$$

That is, correlations decay exponentially, with $1/e$ correlation time

$$\tau = -\frac{1}{\ln \kappa}. \quad (14)$$

Thus, the correlation time increases monotonically from 0 to ∞ as the correlation strength κ is increased from 0 to 1. (In the spatial fluctuator model, the correlation length λ increases monotonically from 0 to ∞ as κ is increased from 0 to 1.)

Even though $s_t = T^t s_0 = s_0$ for all t, in the course of each physical run of a quantum circuit involving m circuit layers, a nontrivial *realization* $R \in \{(1, 0)^{\mathrm{T}}, (0, 1)^{\mathrm{T}}\}^{m+1}$ of fluctuator states is sampled. Equivalently, we can say that, during each run, a realization bit string $b \in \{0, 1\}^{m+1}$ of outcomes is obtained, where the tth entry of b, $b_t \in \{0, 1\}$, denotes the outcome of a trial of the random variable B_t. The realization b occurs with probability

$$p_b = [\delta_{b_0,0}(1 - p) + \delta_{b_0,1}p] \prod_{t=0}^{m-1} T_{b_{t+1}b_t}, \quad (15)$$

using p_b as shorthand notation for $p(B_0 = b_0, B_1 = b_1, \ldots)$. With the above expression, it can be seen with an explicit calculation that, also from the realization viewpoint, the marginalized, time-local probability that the fluctuator is excited at time t is

$$p(B_t = 1) = \sum_{\{b : b_t = 1\}} p_b = p, \quad (16)$$

independent of t and κ.

B. Circuit fidelity

Consider the circuit fidelity $F = \langle \psi | \rho | \psi \rangle$ between the noiseless output state $|\psi\rangle\langle\psi|$ of a VQA's quantum circuit and the noisy output state ρ of that same quantum circuit. If we assume that a single error operator V (Figs. 1 and 2) occurring during the circuit leads to an output state that is orthogonal to the noiseless output state of that circuit, we have in the temporal fluctuator model that $F = (p_{0^m})^n$. Here, p_{0^m} is the probability that no error occurs during the m layers of the circuit, and n is the number of qubits. The probability p_{0^m} is computed straightforwardly with Eq. (15) [53]. For uncorrelated errors, we have $F = (1 - p)^{mn}$, whereas for fully correlated errors, we have $F = (1 - p)^n$. Thus, one may expect an increasing circuit fidelity with increasing correlation strength, and therefore an *improvement* of the performance of VQAs with increasing correlation strength.

The corresponding formulas for the circuit fidelity F in the spatial fluctuator model are obtained by interchanging m and n. Thus, in the case of strong correlations and deep circuits ($m > n$), we expect a lower circuit fidelity in the spatial fluctuator model than in the temporal fluctuator model. Therefore, in the same case, we may expect a worse performance of VQAs in the spatial fluctuator model than in the temporal fluctuator model. Therefore, in the same case, we may expect a worse performance of VQAs in the spatial fluctuator model as compared with in the temporal fluctuator model. Because $mn > m, n$, we expect an even worse performance in the case of fully uncorrelated errors.

However, since one error operator V may cause an output state that is not exactly orthogonal to the noiseless output state, the above discussion merely puts a lower bound on the circuit fidelity. Furthermore, even if the circuit fidelity is exactly known, it merely puts a bound on the AR, which may be loose [54]. Additionally, it is not clear *a priori* how the variational parameters may change due to the noise and the correlations thereof. Finally, because of the breakdown of quantum error correction in the presence of strongly correlated noise, one might expect a *decrease* in the performance of VQAs with increasing correlation strength. In the next section, we therefore derive analytical expressions for the effects of the fluctuator error models on the AR. These are valid in the regime of small error probability and do not account for any possible noise-induced change in the optimal variational parameters. In the section thereafter, we show numerical results that are valid outside the regime of small error probability and account for a possible noise-induced change in optimal variational parameters.

C. Expectation values

Directly after the fluctuator has transitioned from b_{t-1} to b_t, the unitary error operator V is applied to the associated qubit, conditioned on the state of the fluctuator. The transition and the controlled unitary do not change the reduced state of the fluctuator, but correlations are built up by the controlled unitary nevertheless. At the end of the circuit, all fluctuators are traced out. Let $U_1, \ldots, U_m$ be the layers of a quantum circuit (where U_1 denotes the layer that is applied first) and let ρ be its initial state. In QAOA, these layers will depend on the variational parameters $\boldsymbol{\gamma}, \boldsymbol{\beta}$, but in this section, their

dependence on these parameters is mostly suppressed. The effect of a single fluctuator f in the temporal fluctuator model, at the end of the circuit and after tracing out the fluctuator, is described by the noisy circuit

$$\tilde{\mathcal{U}}_f(\rho_0) = \sum_{b^f} p_{b^f} \, \tilde{U}_{b^f} \rho_0 \, \tilde{U}_{b^f}^{\dagger}, \quad \tilde{U}_{b^f} = \prod_{t=m}^{1} V_{b_t^f} U_t, \quad (17)$$

with p_{b^f} as in Eq. (15) (after substituting $b \to b^f$), $\tilde{U}_{b^f}$ being the circuit in case of noise realization b^f, and $\rho_0 = (|+\rangle\langle+|)^{\otimes n}$ being the initial state. Here, $V_{b_t^f}$ is an operator acting immediately after time t and immediately before the gates U_{t+1} on the qubit associated with f, with $V_0 \equiv \mathbb{1}$, $V_1 \equiv V$. [In Eq. (17), the product ordering is such that U_1 appears on the right. When $\tilde{U}_{b^f}$ is applied to a state, the first operator that acts on that state is U_1.]

Because the n realization probability distributions $\{p_{b^f}\}_f$ are independent and identically distributed, their combined effect on the output state is

$$\tilde{\mathcal{U}}(\rho_0) = \sum_{b} p_b \tilde{U}_b \, \rho_0 \, \tilde{U}_b^{\dagger}, \quad (18)$$

with

$$p_b = \prod_{f=n}^{1}(p_{b^f}), \quad \tilde{U}_b = \prod_{t=m}^{1}(V_{b_t} U_t), \quad V_{b_t} = \prod_{f=n}^{1}(V_{b_t^f}). \quad (19)$$

In the spatial fluctuator model, we again have Eq. (18), but with

$$p_b = \prod_{t=m}^{1}(p_{b^t}), \quad \tilde{U}_b = \prod_{t=m}^{1}(V_{b^t} U_t), \quad V_{b^t} = \prod_{q=n}^{1}(V_{b_q^t}). \quad (20)$$

In the spatial model, the different fluctuators carry the label t. Fluctuator t transitions from the state b_q^t to b_{q+1}^t as it "moves" from qubit q to $q+1$.

Returning to the temporal fluctuator model, we define the susceptibility of the cost function to noise at optimal variational parameters (as found by QAOA) as $\chi = (d\langle H\rangle/dp)_{p=0}$, with $\text{tr}[\rho(\gamma_0^*, \beta_0^*)H] \equiv \langle H\rangle$ for short. Using χ, we may write the first-order approximation to the AR, which we call the *linearized AR*, as

$$AR^{\text{lin}}(p) = AR(0) + p\frac{\chi}{C^*}, \quad (21)$$

where we have by Eq. (17) that

$$\chi = \sum_{b} \frac{dp_b}{dp}\bigg|_{p=0} \text{tr}(\tilde{U}_b \rho_0 \tilde{U}_b^{\dagger} H). \quad (22)$$

In this section, we are mainly interested in χ, for which no approximations are made. The linearized AR, which approximates the AR up to an error $O(p^2)$ as p goes to zero, is introduced for later reference.

Since p_b [Eq. (15)] is some explicit polynomial in p and κ, it is clear that the derivative with respect to p at $p = 0$ can be computed analytically. The resulting expression for χ is derived in Appendix A. The expression becomes especially clear in the limits of no correlation ($\kappa = 0$) and full correlation ($\kappa = 1$),

$$\chi_\kappa^{(a)} = |\mathcal{B}_{1+a\delta_{\kappa 1}}|\{\langle H\rangle^{(1+a\delta_{\kappa 1})} - \langle H\rangle_0\}, \quad (23)$$

where $a = m$ for fully temporally correlated errors, and $a = n$ for fully spatially correlated errors. Here, $\mathcal{B}_\ell$ is the set of realizations b where exactly one fluctuator has exactly one chain of ℓ contiguous excitations (no other excitations in any fluctuator present), $\langle H\rangle^{(\ell)}$ is the expectation value of H given realization b, averaged over all $b \in \mathcal{B}_\ell$, and $\langle H\rangle_0$ is the expectation value of H in the noiseless case.

For fully temporally correlated errors, we have the prefactor $|\mathcal{B}_{1+m}| = n$, for fully spatially correlated errors, $|\mathcal{B}_{1+n}| = m$, and for fully uncorrelated errors $|\mathcal{B}_1| = n(m+1)$. So, roughly speaking, there are two effects that determine whether we should expect $\chi_0 > \chi_1$ or $\chi_0 < \chi_1$. On the one hand, one expects that, (i) on average, the detrimental effect of a fully correlated error, consisting of m consecutive errors (n adjacent errors in the case of spatially correlated errors) on $\langle H\rangle$ is larger than the detrimental effect of 1 error on $\langle H\rangle$. That is, one expects $(\langle H\rangle^{(m+1)} - \langle H\rangle_0) > (\langle H\rangle^{(1)} - \langle H\rangle_0)$ and $(\langle H\rangle^{(n+1)} - \langle H\rangle_0) > (\langle H\rangle^{(1)} - \langle H\rangle_0)$. On the other hand, (ii) there are fewer ways in which a fully correlated error can happen: a fully temporally correlated error can happen in n different ways and a fully spatially correlated error can happen in m different ways, whereas a single uncorrelated error can happen in nm ways. [The latter still leads to a prefactor of $|\mathcal{B}_1| = n(m+1)$ because of the way $\langle H\rangle^{(1)}$ is defined.] If the effect (ii) is dominant, we expect that for deep circuits ($m > n$), the AR is highest with fully temporally correlated errors, followed by the AR with fully spatially correlated errors, followed by the fully uncorrelated errors. This order of ARs would be consistent with the expectations that follow from the fidelity arguments in the previous section.

IV. NUMERICAL METHODS AND RESULTS

In this section, we present numerical data on the SWAP-network implementation of QAOA (Fig. 3) under the influence of the temporal and spatial fluctuator models. As the error operator V (Figs. 1 and 2), we choose the bit-phase-flip error or Pauli-Y error. This choice is motivated by the fact that Y is the only nontrivial single-qubit Pauli operator that commutes with neither RZZ' nor RX. To exclude the possibility that an increase in AR as a function of κ is due to $Y^2 = \mathbb{1}$, which may occur, for example, due to Y operators being inserted at $t = 1$ and $t = 2$ on the first qubit, we include an interaction between the fluctuator and its associated qubit only after the qubit was acted on by a RZZ' or RX gate. [Since, in SK models, $\omega_i = 0$, the entire circuit (excluding measurement) consist of RZZ' and RX gates, as in Fig. 3.]

Simulations were performed for 16 random SK instances on $n = 6$ qubits, with $r = 3$ cycles (leading to $m = 21$ circuit layers) and various values of p and κ. For each data point; that is, for each combination of SK instance, type of fluctuator model, p, and κ, a basin-hopping routine [55] was run 32 times to heuristically optimize the cost function landscape $\tilde{C}$. The meta-parameters of the routine were the default values as per SciPy [56], but with four iterations and all initial parameters γ, β chosen at random in the interval $[-0.5 \times 10^{-3}, 0.5 \times 10^{-3}]$. The randomness of the optimization routine on the one hand and the regularity of the outcomes on

the other indicates that the optimization routine consistently found the global optimum, as is discussed in detail in the Supplemental Material [57].

The mixed state ρ' at the end of the quantum circuit was obtained by full density-matrix simulation of all qubits and fluctuators, tracing out all fluctuators before computing the cost function $\tilde{C} = \mathrm{tr}(\rho'H)$ (dependence on variational parameters suppressed). With the full density matrix at hand, the computation of the cost function at given p, κ and variational parameters is free of shot noise and thus essentially exact. The simulator itself was obtained by extending the simulator used in Ref. [58] with the functionality of density-matrix simulation of mixed quantum-classical registers. All code used to generate the results in this work, including the code of the simulator, is available as Supplemental Material [57].

We find qualitatively equal results for all 16 random problem instances. In this section, we report on the results for one typical instance. Details on its typicality and on the other 15 instances are found in Appendix C and the Supplemental Material [57].

We first present the numerical results regarding the noise-adaptivity, followed by our results on the dependence of the AR on the correlation strength κ. Figure 4 shows the AR as a function of p at fixed $\kappa \in \{0, 1\}$ for the typical instance, obtained under the temporal and spatial fluctuator models. Increasing p away from zero, we observe a linear decrease of the AR, with excellent agreement between the AR, the noise-unaware AR, and the linearized AR. In this sense, no robustness of QAOA against errors of the three types is observed around $p = 0$.

For uncorrelated noise, a divergence between the AR and the noise-unaware AR is observed at $p_{\mathrm{crit}}^{\mathrm{uncor}} = (3.9 \pm 0.1) \times 10^{-3}$. This coincides with an abrupt jump in the otherwise slowly varying optimal parameters (see Supplemental Material [57]). This shows that, as p is increased away from zero, initially no other local minimum of the cost function becomes available and that the location of the initial minimum is approximately constant. For $p > p_{\mathrm{crit}}^{\mathrm{uncor}}$, for the first time, a remote lower local minimum becomes available. This new local minimum remains the lowest local minimum as found by QAOA until $p = 0.015 \pm 0.005$. Similar first critical points, not visible in the plot because of the scale, occur at $p_{\mathrm{crit}}^{\mathrm{temp}} = 0.077 \pm 0.001$ for fully temporally correlated noise, and at $p_{\mathrm{crit}}^{\mathrm{spat}} = 0.013 \pm 0.001$ for fully spatially correlated noise. For noise that would break the symmetry of the cost function landscape, a divergence between the noise-aware AR and the noise-unaware AR would already be expected at $p = 0$ [24].

An increase of the noise-aware AR in the case of fully spatially correlated errors is seen after $p \approx 0.2$. This effect may arise because the noise creates a mixture of states, some of which may have a favorable cost. Then, increasing the error probability may in some cases lead to an increase in the AR. This effect is perhaps best illustrated by a simple example. Consider a trivial version of QAOA, with a single qubit in the initial state $|+\rangle$, no variational circuit, and Hamiltonian $H = Z$, under the effect of the error channel $\rho \mapsto (1 - p)\rho + RY(\pi/2)\,\rho\,RY^\dagger(\pi/2)$. Then, with $RY(\pi/2)|+\rangle = |1\rangle$,
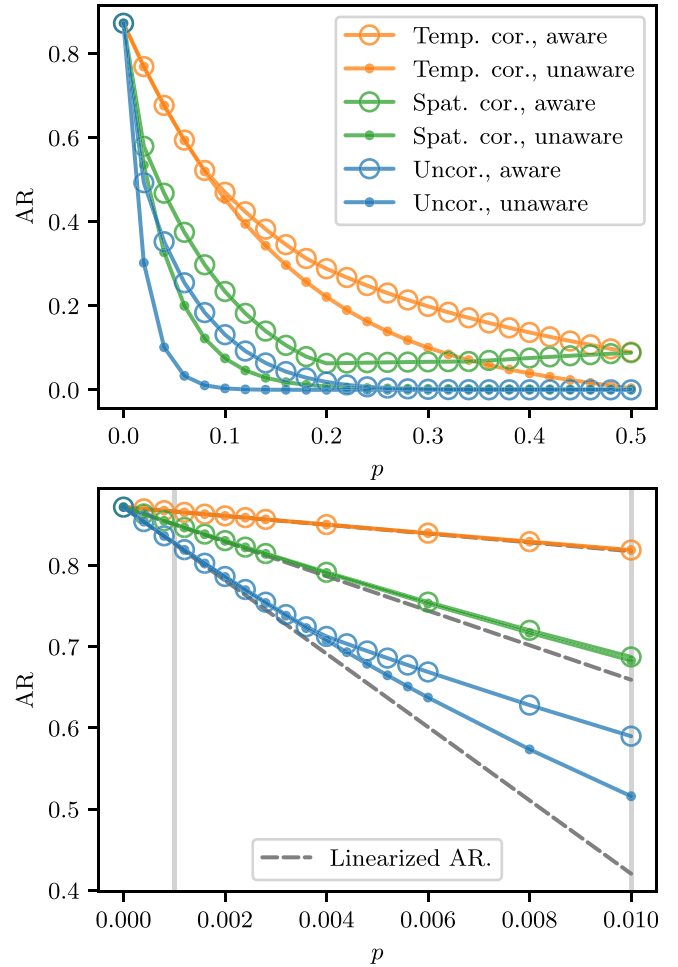

FIG. 4. The approximation ratio (AR), obtained by the noisy SWAP-network implementation of QAOA for a typical problem instance (see Appendix C), for $0 \leqslant p \leqslant 0.5$ (top) and $0 \leqslant p \leqslant 0.01$ (bottom), at fixed correlation strength κ. Results are shown for fully temporally correlated errors (orange), fully spatially correlated errors (green), and fully uncorrelated errors (blue). Open circles show the noise-aware AR [Eq. (6)], whereas the filled circles show the noise-unaware AR [Eq. (7)]. The linearized AR [Eq. (21)] is displayed as a dashed gray line for visibility, irrespective of the noise model and correlation strength. Light gray lines at $p \in \{0.001, 0.01\}$ are for later reference. Error bars are absent because of the use of full density-matrix simulation. The AR in the case of random guessing is exactly zero (Sec. II).

the cost function "landscape" after the error channel is $\tilde{C} = -p$. That is, in this example, $AR(p) = p$.

We now discuss the numerical data and results on the dependence of the AR on the correlation strength. Our conclusion that error correlations can be beneficial for VQAs such as QAOA is based on these data, together with the analytical results of Sec. III C and the additional data in the Supplemental Material [57]. In addition to data on uncorrelated errors, Fig. 4 shows data on the extreme cases of fully temporally and fully spatially correlated errors. Out of the three types of correlation (uncorrelated, temporal, spatial), the AR is highest for fully temporally correlated errors. Second in AR comes the fully spatially correlated errors, followed by the AR for
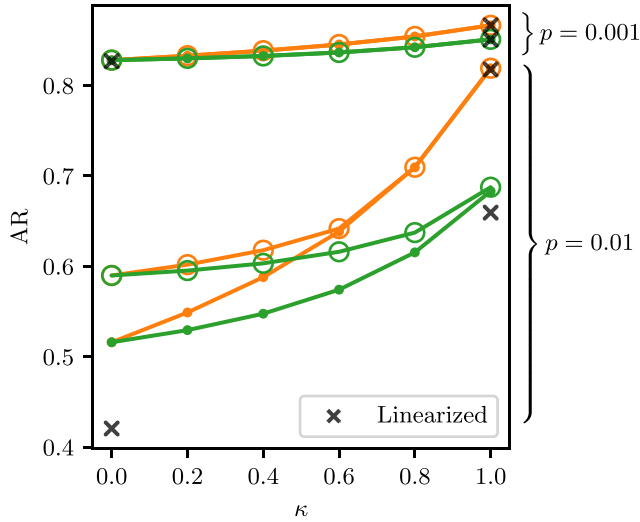
FIG. 5. The approximation ratio (AR) for a typical problem instance (see Appendix C), as a function of the correlation strength κ, at fixed $p = 0.001$ (upper lines and crosses) and $p = 0.01$ (lower lines and crosses). Note that the correlation time τ and the correlation length λ are monotonically increasing functions of κ [Eq. (14)]. At $\kappa = 0$, errors are fully uncorrelated, whereas at $\kappa = 1$, errors are fully correlated. The values for p correspond to the vertical lines in Fig. 4. Definitions are as in Fig. 4, except for the linearized AR, which is now displayed with crosses. At $\kappa = 0$, the temporal fluctuator model is equivalent to the spatial fluctuator model, and hence only two black crosses are shown at $\kappa = 0$, the upper one showing the linearized AR at $p = 0.001$, and the lower one showing the same at $p = 0.01$. For visibility, at $\kappa = 1$, crosses are black irrespective of the type of fluctuator model and p. From top to bottom, they describe the linearized AR at $p = 0.001$ for the temporal fluctuator model and spatial fluctuator model, followed by the linearized AR for the temporal fluctuator model and spatial fluctuator model at $p = 0.01$.

uncorrelated errors. This shows that, for the cases studied here, effect (ii) in Sec. III C is dominant for small p. The order of the ARs remains unchanged for all p considered, indicating that the same effect plays a role for all p [59].

After having discussed these extreme cases, we now take a closer look at the effects of partially correlated errors. Figure 5 shows the AR as a function of κ at fixed $p \in \{0.001, 0.01\}$ for both fluctuator models. These values for p are chosen because typical two-qubit error probabilities roughly fall within the range $0.001 \lesssim p \lesssim 0.01$ [60–62]. The data suggest that the noise-aware AR is highest for temporally correlated errors, and that the increase in the AR as a function of κ is monotonic for all fluctuator models and p considered. At $p = 0.001$, excellent agreement between the noise-aware AR, the noise-unaware AR, and the linearized AR is shown for all $0 \leqslant \kappa \leqslant 1$ considered. At $p = 0.01$, $\kappa = 0$, the agreement between the AR, the noise-unaware AR, and the linearized AR has been broken, as was already apparent from Fig. 4. In Fig. 5, we see how the first critical p depends on κ; $p = 0.001$ is below $p_{\mathrm{crit}}^{\kappa}$ for both fluctuator models and any $0 \leqslant \kappa \leqslant 1$. In the spatial fluctuator model, $p = 0.01$ is *above* $p_{\mathrm{crit}}^{\mathrm{spat},\kappa}$ for all $0 \leqslant \kappa \leqslant 1$. In the temporal fluctuator model, $p = 0.01$ is below $p_{\mathrm{crit}}^{\mathrm{temp},\kappa}$ for κ below roughly 0.7, and above $p_{\mathrm{crit}}^{\mathrm{temp},\kappa}$ for κ above roughly 0.7.

V. DISCUSSION

We studied the performance of QAOA under temporally and spatially correlated errors using a physically inspired toy error model. In the model for temporally correlated errors, each qubit interacts with one independent classical fluctuator during a quantum computation. In the model for spatially correlated errors, after each gate time, all qubits interact with a single fluctuator that is reset before the next gate time. As the fluctuator moves through time (temporal model) or space (spatial model), it undergoes an internal time evolution described by a Markov process. Using the Markov formulation, we showed that the spacetime-local marginalized error probability p is independent of the correlation strength of the Markov chain, and how the latter can be varied from zero to infinity.

We showed that, to first order in the local error probability p, the effect of the fluctuator error models on QAOA's cost function landscape has two competing factors. On the one hand, (i) the detrimental effect of a single correlated error may be worse than the effect of a single uncorrelated error, but on the other, (ii) there are far fewer ways in which a correlated error may act. For example, a fully temporally correlated error (that is, the error operation V, defined in Figs. 1 and 2, is inserted after each gate time) during a circuit acting on n qubits can happen in n ways, a fully spatially correlated error (i.e., an error operation V is inserted at every qubit) during a circuit of depth m can happen in only m ways, whereas a fully uncorrelated error (i.e., a single error operation V is inserted anywhere in the circuit) can happen in nm ways.

We numerically simulated QAOA for 16 random SK problem instances on 6 qubits. For all instances, we observed an increase in the performance of QAOA with correlation strength at all p, indicating that effect (ii) is dominant [63]. As a separate effect, in all instances and at all correlation strengths, we observed the existence of critical local error probabilities after which the noise-aware approximation ratio is higher than the noise-unaware approximation ratio. This shows QAOA can in some cases adapt to the effects of uncorrelated and correlated errors, given that error probabilities are *above* a threshold. It remains an open question how this threshold behaves as the number of qubits is increased to the numbers required for a practical quantum advantage. The noise resilience of QAOA may be seriously limited if the noise-adaptivity threshold increases with n.

Our work does not show that correlated errors are beneficial to QAOA; adding correlated errors *on top* of local errors will almost certainly result in a reduction in the performance of QAOA. What our work indicates is that it is not correlation in itself that has a negative effect on QAOA. To improve current hardware to the point that it can demonstrate a useful quantum advantage, a reduction of both uncorrelated and correlated errors remains necessary.

The error model used in this work is minimalistic, but we do not expect that the qualitative results of our work depend on the details of more complicated or realistic noise models. A first limitation of our noise model is that it is ultimately a stochastic unitary error model, where unitary error operations are inserted into the circuit according to some (correlated) probability distribution [Eq. (18)]. Not all quantum channels are of this type, such as the amplitude damping channel,

but they can be transformed into an effective stochastic error model (specifically, a stochastic Pauli error model) by the process of Pauli twirling [35,64]. As a result, if QAOA is run on a real quantum device with added Pauli twirling operations, the physical noise experienced during the quantum computation will effectively be formed by stochastic Pauli errors, thereby also reinstating the symmetries of the SK QAOA cost function landscape derived in Appendix B.

A second limitation of our noise model is that all stochastic error unitaries inserted into the circuit are equal. Although allowing, e.g., gate-dependent errors would result in a more realistic noise model, we expect that our results would not be qualitatively affected. This is because we expect the states during the ansatz circuit to be sufficiently spread out over the Hilbert space, and to vary enough from circuit layer to circuit layer, that the choice of error operation V is irrelevant as long as it does not fully commute with the gates in the ansatz circuit. If, on the other hand, the error operator were to commute with, for example, the RZZ gates, it would effectively lead to one layer of error operations per circuit cycle, coarsening the error model.

A final limitation of our error model is that it only considers error correlations with an exponentially decaying correlation function [Eq. (13)]. This does not include scenarios in which the correlations decay as a power law. However, it is important to note that if our model were to be expanded to accommodate power-law correlations, it would be equivalent to the current model in the cases of uncorrelated and fully correlated errors, therefore leaving Fig. 4 unaltered. The same holds for errors with a $1/f$ spectral density. Additionally, $1/f$ noise has a correlation function proportional to $\mathrm{Ei}(-f_0|t|) = O(e^{-f_0|t|})$ [45], with f_0 a low-frequency cutoff. Hence, a low-frequency cutoff introduces a timescale that confines correlations to within a finite time span, much like the correlations in our model.

Lines for future work include the study of the generality of our results. It is *a priori* not clear if similar effects hold in more general models of temporally and spatially correlated noise, or if they hold for other VQAs. If they do, the experimental requirements on noise correlation strengths may not be as stringent as previously thought, bringing a useful quantum advantage within closer reach.

Code and data is available as Supplemental Material [57] and at GitHub [65].

APPENDIX A: COST FUNCTION LANDSCAPE NOISE SUSCEPTIBILITY

Here, we analytically compute χ [Eq. (22)], the derivative of the cost function landscape $\tilde{C}$ [Eq. (2)] (at fixed variational parameters) to the local error probability p at $p = 0$. The only assumption on $\tilde{C}$ is that it is obtained using a circuit that is

acted upon by our fluctuator models with noise parameters p and κ (Figs. 1 and 2).

Let us first consider a single fluctuator f. Denote the noise realizations of that fluctuator by the bit string b^f. Computing the derivative of Eq. (15), we find

$$\frac{dp_{b^f}}{dp}\bigg|_{p=0} = \sum_{\ell=1}^{m+1} \sum_{i=0}^{m+1-\ell} \delta_{b^f, 0^i 1^\ell 0^{k_{i\ell}}} (1-\kappa)^{2-\delta_{i,0}-\delta_{k_{i\ell},0}} \kappa^{\ell-1} - \delta_{b^f, \mathbf{0}}[m(1-\kappa)+1], \tag{A1}$$

with $k_{i\ell} = m + 1 - i - \ell$ the number of zeros that need to be appended to the bit string $0^i 1^\ell$ to make it a bit string of length $m + 1$.

Moving to n fluctuators, we have by Eq. (19) that

$$\frac{dp_b}{dp}\bigg|_{p=0} = \sum_{f=1}^{n} \delta_{b^1 \dots b^{f-1}, \mathbf{0}} \left(\frac{dp_{b^f}}{dp}\right)_{p=0} \delta_{b^{f+1} \dots b^n, \mathbf{0}}. \tag{A2}$$

Then, with the definition $\langle H \rangle_b = \mathrm{tr}(\tilde{U}_b \rho_0 \tilde{U}_b^\dagger H)$, we have by Eq. (22), that

$$\chi = \sum_{b,f,\ell,i} \tilde{\delta}_{b,f,\ell,i}(1-\kappa)^{2-\delta_{i,0}-\delta_{k_{i},0}} \kappa^{\ell-1} \langle H \rangle_b - n[m(1-\kappa)+1]\langle H \rangle_{\mathbf{0}}, \tag{A3}$$

with $\tilde{\delta}_{b,f,\ell,i} = \delta_{b^1 \dots b^{f-1}, \mathbf{0}} \delta_{b^f, 0^i 1^\ell 0^{k_{i\ell}}} \delta_{b^{f+1} \dots b^n, \mathbf{0}}$, and where the sum is over all b, i, j, f, in the ranges as described before. Note $\mathbf{0} = 00 \dots 0$ has varying dimension depending on context. We can write the previous expression for χ more schematically as

$$\chi = \sum_b [(1-\kappa)^{h(b)} \kappa^{\ell(b)-1} \langle H \rangle_b] - n[m(1-\kappa)+1]\langle H \rangle_{\mathbf{0}}, \tag{A4}$$

where the sum is over all realizations b where exactly one fluctuator has exactly one chain of consecutive excitations (and no other excitations), $h(b) \in \{0, 1, 2\}$ is the number of transitions $0 \leftrightarrow 1$ in b, and $\ell(b) \geqslant 1$ is the length of the chain.

The average contribution of chains of errors of length ℓ to χ schematically reads $\langle H \rangle^{(\ell)} \equiv \frac{1}{|\mathcal{B}_\ell|} \sum_{b \in \mathcal{B}_\ell} (1-\kappa)^{h(b)} \langle H \rangle_b$, where the set $\mathcal{B}_\ell$ contains those realizations b where exactly one fluctuator has exactly one chain of excitations of length ℓ, and where we have absorbed the boundary effects $(1-\kappa)^{h(b)}$ into $\langle H \rangle^{(\ell)}$. With this definition, the susceptibility becomes

$$\chi = \sum_{\ell=1}^{m+1} (|\mathcal{B}_\ell| \kappa^{\ell-1} \langle H \rangle^{(\ell)}) - n[m(1-\kappa)+1]\langle H \rangle_{\mathbf{0}}. \tag{A5}$$

APPENDIX B: COST FUNCTION LANDSCAPE SYMMETRIES FOR SHERRINGTON-KIRKPATRICK PROBLEM INSTANCES

Here, we prove that a certain set of transformations of the parameters of the QAOA SK cost function landscape generates a symmetry group of the cost function landscape. Note that the SK Hamiltonian is given by Eq. (1) with $\omega_{ij} \in$

$\{1, -1\}$, $\omega_i = 0$ and that hence the Hamiltonian itself possesses a global $\mathbb{Z}_2$ symmetry. Additionally, because $\omega_i = 0$, the ansatz circuit [Eq. (3), Fig. 3] does not contain RZ gates. The numerical data available in the Supplemental Material [57] suggest that the transformations derived in the current appendix generate the entire symmetry group of the cost function landscape. The symmetries are not broken under a general class of noise models that we call the spatiotemporally correlated Pauli noise channels. The main application of the results in this section is that they allow for the comparison of optimal variational parameters modulo the symmetries of the cost function landscape.

1. Symmetry group

The class of Pauli channels contains all channels of the form

$$\Lambda(\rho) = \sum_a p_a P_a \rho P_a^\dagger, \tag{B1}$$

with p_a being a probability distribution, and $\{P_a\}_a$ being tensor products of the Pauli operators $\{\mathbb{1}, X, Y, Z\}$, also referred to as Pauli words. Note that Pauli channels have a stochastic interpretation: with probability p_a, the Pauli word P_a acts on ρ. We generalize this to with multiple layers and insert Pauli words according to a correlated probability distribution.

Definition B.1. (Spatiotemporally correlated Pauli channels) Let $U = U_g \cdots U_1$ be a noiseless quantum circuit consisting of the unitaries $\{U_i\}_i$. The spatiotemporally correlated Pauli channels are of the form

$$\Lambda(\rho) = \sum_a p_a \tilde{U}_a \rho \tilde{U}_a^\dagger, \tag{B2}$$

with

$$\tilde{U}_a = \prod_{i=g}^1 (P_{a_i} U_i) \tag{B3}$$

being the circuit realization in case of noise realization a.

Spatiotemporally correlated Pauli channels effectively insert Pauli operations into the circuit stochastically and can hence be considered to be a form of stochastic unitary error, where all unitaries are Pauli operators and where the probability distribution may be correlated in space and time. The fluctuator models of the main text are spatiotemporally correlated Pauli channels if the error operation V is a Pauli operator. Note that if a set of transformations generates a symmetry group of the cost function landscape, also the inverses and any composition of those transformations form symmetries of the cost function landscape.

Lemma B.1. Any n-qubit QAOA SK cost function landscape $\tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$, obtained using any number of SWAP gates while the ansatz circuit was acted upon by any spatiotemporally correlated Pauli channel, is invariant under the group generated by the following transformations:

  (1) translation of any γ_k by 2π;
  (2) translation of any β_k by 1π;
  (3) negation of any β_k (i.e., $\beta_k \mapsto -\beta_k$) and simultaneous translation of γ_k and γ_{k+1} by π (for the last β parameter; that

is, the edge case $k = r$, only $\gamma_k = \gamma_r$ needs to be translated by π);
  (4) simultaneous negation of all parameters.

Proof. Let us first assume the QAOA circuit is implemented in the absence of noise. In the following, we consider two states or two unitaries equal if they are equal up to a global phase. The first generator follows directly from the periodicity of $RZZ(\gamma)$ with period 2π. For the second generator, note that $RX(\beta \pm \pi) = X\,RX(\beta)$. Thus, sending $\beta_k \mapsto \beta_k \pm \pi$ has the same effect on the ansatz state as adding a layer of X gates immediately before the layer of RX gates of the cycle k. Now, note that the layer of X gates commutes with all RZZ gates, any SWAP gates, and all layers of RX gates at all cycles of the ansatz circuit. Thus, we may commute the layer of X gates all the way to the beginning of the circuit and act with it on the initial state, resulting in $X^{\otimes n}|+\rangle^{\otimes n} = |+\rangle^{\otimes n}$. Thus, the ansatz state $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$ is invariant under $\beta_k \mapsto \beta_k \pm \pi$ for any k, and consequently $\tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is invariant under these transformations.

For the third generator, note that $RZZ[\pm'(\gamma \pm'' \pi)] = (Z \otimes Z)\,RZZ[\pm'(\gamma \pm'' \pi)]$. (The signs $\pm'$ and $\pm''$ are independent.) Thus, sending $\gamma_k \mapsto \gamma_k \pm' \pi$ has the same effect as adding $Z \otimes Z$ to the circuit immediately after each RZZ gate in the Hamiltonian stage of the cycle k. Note that $Z_i Z_j$ commutes with $RZZ_{kl}(x)$ for all x, i, j, k, l and that any SWAP gate can be seen as permuting qubit indices, not changing the number of times a qubit is acted on by an RZZ gate. We may now commute all excess Z gates that arose from the shift $\gamma_k \mapsto \gamma_k \pm'' \pi$ to immediately before the layer of RX gates of cycle k. At this layer, each qubit is now acted on by an odd number of Z gates, which is equal to the situation where each qubit is acted on by a single Z gate. We can do the same for the cycle $k + 1$ of the ansatz circuit, sending $\gamma_k \mapsto \gamma_k \pm'' \pi$, but now moving the resulting Z gates to immediately before the layer of RX gates of the cycle k. Using $Z_i RX_i(\beta_k)Z_i = RX_i(-\beta_k)$ on all qubits proves the standard case of generator 3. For the edge case $k = r$, sending $\gamma_r \mapsto \gamma_r \pm'' \pi$ results in a layer of Z gates immediately before the last layer of RX gates. Commuting this layer through the last layer of RX gates sends β_r to $-\beta_r$. Noting that the layer of Z gates commutes through (any term of) H, and thus that $\langle \boldsymbol{\gamma}, \boldsymbol{\beta}|Z^{\otimes n}HZ^{\otimes n}|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle = \langle \boldsymbol{\gamma}, \boldsymbol{\beta}|H|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle = \tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$, completes the proof of generator 3 (in the noiseless case).

The fourth generator follows because (in the computational basis) (*i*) any gate in the ansatz is generated by a real Hamiltonian, (*ii*) the Hamiltonian used in QAOA is real, and (*iii*), the initial state is real. Both $U(\theta) = RX(\theta)$ and $U(\theta) = RZZ(\theta)$ are generated by real Hamiltonians [i.e., statement (*i*) holds]. Thus, for both operators, $U^T(\theta) = U^\dagger(-\theta)$ and, likewise, $U^{\dagger T}(\theta) = U(-\theta)$. From (*ii*) and (*iii*), it follows that $H^T = H$ and $(|+\rangle^n)^T = \langle +|^n$. Let the sequence of gates in the ansatz be given by $\Pi_{i=g}^1 U_i(\theta_i)$, with g being the number of gates in the ansatz, and note that trivially $x = x^T$ for x a real number. Then,

$$\begin{aligned}
\tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) &= \left(\langle +|^n \left[\Pi_{i=1}^g U_i^\dagger(\theta_i)\right] H \left[\Pi_{i=g}^1 U_i(\theta_i)\right]|+\rangle^n\right)^T \\
&= \langle +|^n \left[\Pi_{i=1}^g U_i^\dagger(-\theta_i)\right] H \left[\Pi_{i=g}^1 U_i(-\theta_i)\right]|+\rangle^n \\
&= \tilde{C}(-\boldsymbol{\gamma}, -\boldsymbol{\beta}). \tag{B4}
\end{aligned}$$

This completes the proof of Lemma B.1 in the noiseless case.

In case the ansatz circuit is acted upon by a spatiotemporally correlated Pauli error channel,

$$\tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_a p_a \langle +|^{\otimes n} \tilde{U}_a^\dagger(\boldsymbol{\gamma}, \boldsymbol{\beta}) H \, \tilde{U}_a(\boldsymbol{\gamma}, \boldsymbol{\beta}) |+\rangle^{\otimes n}$$

$$:= \sum_a p_a \tilde{C}_a(\boldsymbol{\gamma}, \boldsymbol{\beta}). \tag{B5}$$

So, to show that $\tilde{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is invariant under the transformations generated by generators 1–4, it suffices to show that $\tilde{C}_a(\boldsymbol{\gamma}, \boldsymbol{\beta})$, the cost function in case of realization a, is invariant under those transformations for all noise realizations a.

Transformation 1 trivially leaves $\tilde{C}_a(\boldsymbol{\gamma}, \boldsymbol{\beta})$ invariant under any spatiotemporally correlated Pauli error channel. To show that the transformations 2 and 3 leave $\tilde{C}_a(\boldsymbol{\gamma}, \boldsymbol{\beta})$ invariant under the same type of channel, consider $U_a(\boldsymbol{\gamma}, \boldsymbol{\beta})$, the circuit in case of noise realization a. Note that any Pauli operator either commutes or anticommutes with both the layer of X gates arising from any the parameter shift $RX(\beta \pm \pi) = X \, RX(\beta)$, and any $Z \otimes Z$ operators originating from $RZZ[\pm'(\gamma \pm'' \pi)] = Z \otimes Z \, RZZ[\pm'(\gamma \pm'' \pi)]$. Hence, we can use the noiseless proofs of generators 2 and 3 to show that $C_a(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is invariant for all a if we make the additional observation that any overall factors of -1, arising from anticommutativity, have no effect on the expectation value $C_a(\boldsymbol{\gamma}, \boldsymbol{\beta})$.

To show that generator 4 leaves $\tilde{C}_a(\boldsymbol{\gamma}, \boldsymbol{\beta})$ invariant, note that $RZZ_{ij} = RZZ_{ji}$, and that $RZZ(\theta)P_i = P_i RZZ_{ij}(-\theta)$, if $P_i = X_i$ or $P_i = Y_i$, and $RZZ_{ij}(\theta)P_i = P_i RZZ_{ij}(\theta)$ if $P_i = Z$. Likewise, $RX(\theta)P = P \, RX(-\theta)$ if $P = Y$ or $P = Z$, and $RX(\theta)P = P \, RX(\theta)$ if $P = X$. Commuting all Pauli operators arising from an error realization to the beginning of the circuit, we obtain $\tilde{U}_a(\boldsymbol{\gamma}, \boldsymbol{\beta}) = U_{\mathbf{0}}(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\beta}})P$, with P some tensor product of Pauli operators, $\tilde{U}_{\mathbf{0}}$ the noiseless circuit realization, and where $\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\beta}}$ can be determined explicitly using the aforementioned commutation relations. Thus, $U_a(\boldsymbol{\gamma}, \boldsymbol{\beta})|+\rangle^{\otimes n} = U_{\mathbf{0}}(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\beta}})|\psi\rangle$, with $|\psi\rangle$ a state with real entries. Hence, conditions (i)–(iii) used in the noiseless proof of generator 4 are satisfied, and therefore $C_a(\boldsymbol{\gamma}, \boldsymbol{\beta}) = C_a(-\boldsymbol{\gamma}, -\boldsymbol{\beta})$. ∎

2. Parameter representatives

Due to the symmetries of the SK QAOA cost function landscape, two unequal sets of parameters $\Phi = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ and $\Phi' = (\boldsymbol{\gamma}', \boldsymbol{\beta}')$ may be deemed equivalent if there exists a group element from the symmetry group of the QAOA SK cost function landscape that maps Φ to Φ'. This defines equivalence classes of sets of parameters. Algorithm 1 maps each set of parameters Φ to a unique representative $\mathcal{A}(\Phi)$ of the equivalence class that set of parameters is in. Thus, the Euclidean distance D between Φ and Φ', modulo the symmetries of the cost function, may be defined as

$$D = \sqrt{\sum_i [\mathcal{A}(\Phi)_i - \mathcal{A}(\Phi')_i]^2}. \tag{B6}$$

This is the procedure used in the Supplemental Material [57] to plot the distance D between the noise-unaware parameters and the noise-aware parameters.

Algorithm 1. The algorithm that removes the redundant freedom that is caused by symmetries of the QAOA SK cost function landscape. It does so by mapping each set of parameters to a unique class representative.

---

input : $(\boldsymbol{\gamma}, \boldsymbol{\beta})$
output: unique representative of the
  parameter class that $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ belongs to

Function $compress[(\boldsymbol{\gamma}, \boldsymbol{\beta})]$ **is**
  for $i \leftarrow 0$ **to** r **do**
    $\gamma_i \leftarrow \gamma_i \bmod 2\pi$
    if $\gamma_i \geq \pi$ **then**
      $\gamma_i \leftarrow \gamma_i - \pi$
      $\beta_i \leftarrow -\beta_i$
      if $i \leq r - 2$ **then**
        $\gamma_{i+1} \leftarrow \gamma_{i+1} - \pi$
      end
    end
    $\beta_i \leftarrow \beta_i \bmod \pi$
  end
  return $(\boldsymbol{\gamma}, \boldsymbol{\beta})$
end

$(\boldsymbol{\gamma}, \boldsymbol{\beta}) \leftarrow compress[(\boldsymbol{\gamma}, \boldsymbol{\beta})]$
$(\boldsymbol{\gamma}', \boldsymbol{\beta}') \leftarrow compress[(-\boldsymbol{\gamma}, -\boldsymbol{\beta})]$

if $|\boldsymbol{\gamma}| < |\boldsymbol{\gamma}'|$ **then**
  return $(\boldsymbol{\gamma}, \boldsymbol{\beta})$
else
  return $(\boldsymbol{\gamma}', \boldsymbol{\beta}')$
end

APPENDIX C: INSTANCE PROPERTIES

The instance reported on in Sec. IV is

$$H = \sum_{i<j} \omega_{ij} Z_i Z_j, \tag{C1}$$

with

$$\omega = \begin{pmatrix} 0 & +1 & -1 & +1 & +1 & -1 \\ 0 & 0 & +1 & -1 & +1 & +1 \\ 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & +1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{C2}$$

We can write the data defining the instance more compactly by gathering and concatenating the factors ± 1 from left to right and top to bottom. If we additionally make the identification $(+1, -1) \mapsto (0, 1)$, we obtain the *instance bit string*

$$\tilde{\omega} = 010010100111110. \tag{C3}$$

In general, the instance bit string of a SK model on n qubits contains $n_{\tilde{\omega}} = n(n-1)/2$ bits and reads $\tilde{\omega} = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} [(1 - \omega_{ij})/2]$. Here, the product acts by concatenation (not multiplication) and the bit string is built up

TABLE I. The random instances and some of their properties. None of the 16 graphs are mutually isomorphic, except for those noted with an asterisk.

Instance bit string	Graph class cardinality	GS manifold	GS energy	Cost fn. landscape class
000100100011110	360	010101	-7	0
001100110101110	180	001001	-9	1
001111110000010	720	011010	-9	1
010011011101010	180	010010	-9	1
001111011010001	720	011011	-9	2
010100101011100	360	010101	-9	2
100010101000100	360	001110	-9	2
110010010101000	720	010010	-9	2
010010100111110*	720	010000, 010110	-7	3
011100011010011	360	010011, 010101	-7	3
100110010001101	720	001001, 001100	-7	3
101100001010110	360	001010, 010001	-7	3
111011011000010*	720	000010, 011010	-7	3
111100111011001	360	000100, 001011	-7	3
011101001011100	180	011001	-11	4
110101111100011	72	000000, 000011, 000101 001100, 010111, 011101	-5	5

from left to right [i.e., if $n = 6$, $\tilde{\omega}_1 = (1 - \omega_{1,2})/2$, $\tilde{\omega}_2 = (1 - \omega_{1,3})/2, \ldots, \tilde{\omega}_{15} = (1 - \omega_{5,6})/2$].

We sampled 16 random SK problem instances (including the one above) on $n = 6$ qubits by sampling 16 bit strings of length 15 uniformly at random. Table I lists the following properties of these instances.

(1) *Instance bit string.*

(2) *Graph class cardinality.* Many of the $2^{n_{\tilde{\omega}}}$ SK Hamiltonians on n qubits are equivalent up to a relabeling of the qubit indices. That is, many of the SK graphs with n nodes, obtained by interpreting the ω_{ij} as entries of an adjacency matrix, are isomorphic. Indeed, for $n = 6$, we found by a brute-force method that the $2^{n_{\tilde{\omega}}} = 32\,768$ SK graphs fall into 156 different equivalence classes, where equivalence is defined by graph isomorphism. The graph class cardinality is the number of distinct graphs in a graph class. The highest occurring class cardinality is 720, and there are 8 classes attaining this cardinality. From the 16 random graphs we generated, only the graphs with instance bit strings 010010100111110 and 111011011000010 are in the same class (i.e., isomorphic). The class they belong to has the highest occurring cardinality of 720.

(3) *Ground-state manifold.* We indicate the ground-state manifold of the SK instances by lists of bit strings. The list of an instance is related to the instance's ground-state manifold GS by GS = $\text{span}(\cup_i \{|i\rangle, |\bar{i}\rangle\})$, where i ranges over all bit strings in the list and $\bar{i}$ indicates the negation of i. It follows from $[X^{\otimes n}, H] = 0$ that $|\bar{i}\rangle$ is in the ground-state manifold if $|i\rangle$ is in the ground-state manifold.

(4) *Ground-state energy.* The energy E_{GS} of any state in the ground-state manifold.

(5) *Cost function landscape class.* Consistent with QAOA's cost function landscape concentration [66], we find

that two nonisomorphic SK instances may have identical cost function landscapes (up to numerical precision) at $n = 6$, $d = 3$, $p = 0$. We consider two SK instances to be in the same cost function landscape class if their cost function landscapes are identical. The 16 random SK instances studied in this work, of which only two are isomorphic, fall into 6 distinct cost function landscape classes.

Cost function landscape equivalence is established numerically by the comparison of cost function landscape values at 64 random points in parameter space. We deem a pair of SK instances equivalent to numerical accuracy if their cost function landscape values differ by less than 10^{-14} in absolute value across all 64 random points. We found that, by this criterion, the set of pairs of instances that lie in the same cost function landscape class is well separated from the set of pairs of instances that do not. (For each pair of instances in the same cost function landscape class, there is no point in parameter space among the 64 sampled points where the pair's cost functions differ by more than 10^{-4}. For each pair of instances *not* in the same cost function landscape class, we found that there is at least one point in parameter space where their cost functions differ by more than unity.) Increasing the error probability p breaks the cost function equivalence. The data presented in the Supplemental Material [57] show that this does not lead to qualitative differences in the AR of instances in the same cost function landscape class.

In the main text, we have shown the data pertaining to the SK instance with instance bit string 010010100111110. We consider this instance typical because it falls into a graph class with the highest possible cardinality and because the qualitative performance of QAOA on this instance is largely equal to the qualitative performance of QAOA on all other sampled instances.

[1] R. P. Feynman, Simulating physics with computers, Int. J. Theor. Phys. **21**, 467 (1982).

[2] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[3] P. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (IEEE Comput. Soc. Press, 1994).

[4] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC '96* (ACM Press, 1996).

[5] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, Phys. Rev. Lett. **103**, 150502 (2009).

[6] D. Suter and G. A. Álvarez, Colloquium: Protecting quantum information against environmental noise, Rev. Mod. Phys. **88**, 041001 (2016).

[7] P. Shor, Fault-tolerant quantum computation, in *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE Comput. Soc. Press, 1996).

[8] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing - STOC '97* (ACM Press, 1997).

[9] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation: Error models and thresholds, Proc. R. Soc. London, Ser. A **454**, 365 (1998).

[10] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. **38**, 1207 (2008).

[11] B. M. Terhal and G. Burkard, Fault-tolerant quantum computation for local non-Markovian noise, Phys. Rev. A **71**, 012336 (2005).

[12] D. Aharonov, A. Kitaev, and J. Preskill, Fault-Tolerant Quantum Computation with Long-Range Correlated Noise, Phys. Rev. Lett. **96**, 050504 (2006).

[13] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 code, Quantum Inf. Comput. **6**, 97 (2006).

[14] J. Preskill, Sufficient condition on noise correlations for scalable quantum computing, Quantum Inf. Comput. **13**, 181 (2013).

[15] B. D. Clader, C. J. Trout, J. P. Barnes, K. Schultz, G. Quiroz, and P. Titum, Impact of correlations and heavy tails on quantum error correction, Phys. Rev. A **103**, 052428 (2021).

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

[17] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[18] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, Quantum computational advantage using photons, Science **370**, 1460 (2020).

[19] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[20] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[21] A. Lucas, Ising formulations of many NP problems, Front. Phys. **2**, 5 (2014).

[22] M. Hodson, B. Ruck, H. Ong, D. Garvin, and S. Dulman, Portfolio rebalancing experiments using the quantum alternating operator ansatz, arXiv:1911.05296.

[23] J. R. Weggemans, A. Urech, A. Rausch, R. Spreeuw, R. Boucherie, F. Schreck, K. Schoutens, J. Minář, and F. Speelman, Solving correlation clustering with QAOA and a Rydberg qudit system: A full-stack approach, Quantum **6**, 687 (2022).

[24] E. Fontana, N. Fitzpatrick, D. M. Ramo, R. Duncan, and I. Rungger, Evaluating the noise resilience of variational quantum algorithms, Phys. Rev. A **104**, 022403 (2021).

[25] C. Xue, Z.-Y. Chen, Y.-C. Wu, and G.-P. Guo, Effects of quantum noise on quantum approximate optimization algorithm, Chin. Phys. Lett. **38**, 030302 (2021).

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

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

[28] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding *et al.*, Scalable Quantum Simulation of Molecular Energies, Phys. Rev. X **6**, 031007 (2016).

[29] 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**, 042308 (2017).

[30] 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**, 011021 (2018).

[31] L. Gentini, A. Cuccoli, S. Pirandola, P. Verrucchi, and L. Banchi, Noise-resilient variational hybrid quantum-classical optimization, Phys. Rev. A **102**, 052414 (2020).

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

[33] E. Farhi and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm, arXiv:1602.07674.

[34] I. Heinz and G. Burkard, Crosstalk analysis for single-qubit and two-qubit gates in spin qubit arrays, Phys. Rev. B **104**, 045420 (2021).

[35] E. v.d. Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors, arXiv:2201.09866.

[36] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, 14-Qubit Entanglement: Creation and Coherence, Phys. Rev. Lett. **106**, 130506 (2011).

[37] C. D. Wilen, S. Abdullah, N. Kurinsky, C. Stanford, L. Cardani, G. d'Imperio, C. Tomei, L. Faoro, L. Ioffe, C. Liu *et al.*, Correlated charge noise and relaxation errors in superconducting qubits, Nature (London) **594**, 369 (2021).

[38] L. Cardani, F. Valenti, N. Casali, G. Catelani, T. Charpentier, M. Clemenza, I. Colantoni, A. Cruciani, G. D'Imperio, L. Gironi

et al., Reducing the impact of radioactivity on quantum circuits in a deep-underground facility, Nat. Commun. **12**, 2733 (2021).

[39] P. Dutta and P. M. Horn, Low-frequency fluctuations in solids: $\frac{1}{f}$ noise, Rev. Mod. Phys. **53**, 497 (1981).

[40] R. H. Koch, J. Clarke, W. M. Goubau, J. M. Martinis, C. M. Pegrum, and D. J. van Harlingen, Flicker $(1/f)$ noise in tunnel junction dc SQUIDS, J. Low Temp. Phys. **51**, 207 (1983).

[41] R. H. Koch, D. P. DiVincenzo, and J. Clarke, Model for $1/f$ Flux Noise in Squids and Qubits, Phys. Rev. Lett. **98**, 267003 (2007).

[42] E. J. Connors, J. Nelson, L. F. Edge, and J. M. Nichol, Charge-noise spectroscopy of Si/SiGe quantum dots via dynamically-decoupled exchange oscillations, Nat. Commun. **13**, 940 (2022).

[43] M. Reed, Entanglement and quantum error correction with superconducting qubits, PhD thesis, Yale university, 2013.

[44] A. V. Khaetskii, D. Loss, and L. Glazman, Electron Spin Decoherence in Quantum Dots due to Interaction with Nuclei, Phys. Rev. Lett. **88**, 186802 (2002).

[45] G. Burkard, Non-Markovian qubit dynamics in the presence of $1/f$ noise, Phys. Rev. B **79**, 125317 (2009).

[46] S. Schlör, J. Lisenfeld, C. Müller, A. Bilmes, A. Schneider, D. P. Pappas, A. V. Ustinov, and M. Weides, Correlating Decoherence in Transmon Qubits: Low Frequency Noise by Single Fluctuators, Phys. Rev. Lett. **123**, 190502 (2019).

[47] J. J. Burnett, A. Bengtsson, M. Scigliuzzo, D. Niepce, M. Kudra, P. Delsing, and J. Bylander, Decoherence benchmarking of superconducting qubits, npj Quantum Inf. **5**, 54 (2019).

[48] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum Simulation of Electronic Structure with Linear Depth and Connectivity, Phys. Rev. Lett. **120**, 110501 (2018).

[49] B. O'Gorman, W. J. Huggins, E. G. Rieffel, and K. B. Whaley, Generalized swap networks for near-term quantum computing, arXiv:1905.05118.

[50] A SWAP gate effectively swaps qubit indices. This has to be accounted for by changing the parameters in subsequent RZZ gates, which depend on the effective qubit labels the RZZ gate is acting on, accordingly. One way of implementing this is by keeping a list of qubit indices, initialized at $x = (1, 2, \ldots, n)$. Every time a SWAP gate acts on qubits i and j, the list is updated with $x_i \leftrightarrow x_j$. Each $RZZ(\omega_{ij}\gamma_k)$ has to be replaced with $RZZ(\omega_{x_ix_j}\gamma_k)$ using the state of the list x at the time the RZZ gate acts. For an even number of cycles, the output state of the SWAP-network implementation is identical to the output of the circuit in Eq. (3). For an uneven number of cycles, the effective order of the qubits is reversed, which can be easily accounted for in the classical postprocessing of the measurement outcomes.

[51] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2010).

[52] R. Serfozo, *Basics of Applied Stochastic Processes* (Springer-Verlag, Berlin Heidelberg, 2009).

[53] Because of the way we have defined the interaction between the fluctuator and the qubit, there are two ways in which no errors can occur; $p_{0^m} = \sum_{b_0} p_{b_0 0^m}$. By Eq. (15), we thus have $p_{0^m} = (1 - p)[1 - p(1 - \kappa)]^m + p(1 - \kappa)(1 - p)[1 - p(1 - \kappa)]^{(m-1)}$.

[54] M. J. S. Beach, R. G. Melko, T. Grover, and T. H. Hsieh, Making trotters sprint: A variational imaginary time ansatz for quantum many-body systems, Phys. Rev. B **100**, 094434 (2019).

[55] D. J. Wales and J. P. K. Doye, Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms, J. Phys. Chem. A **101**, 5111 (1997).

[56] SCIPY 1.5.2 documentation (accessed July, 2022).

[57] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.107.042426 for the code and data used to generate the results in this work; see also [65].

[58] J. Kattemölle and J. van Wezel, Variational quantum eigensolver for the Heisenberg antiferromagnet on the kagome lattice, Phys. Rev. B **106**, 214429 (2022).

[59] A single exception occurs for one instance at $p \geqslant 0.4$. Here, the noise-aware AR in the case of fully spatially correlated errors becomes higher than the noise-aware AR in the case of fully temporally correlated errors. The noise-aware AR remains higher for correlated errors. See the Supplemental Material [57] for details.

[60] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, High-Fidelity Quantum Logic Gates using Trapped-Ion Hyperfine Qubits, Phys. Rev. Lett. **117**, 060504 (2016).

[61] X. Xue, M. Russ, N. Samkharadze, B. Undseth, A. Sammak, G. Scappucci, and L. M. Vandersypen, Quantum logic with spin qubits crossing the surface code threshold, Nature (London) **601**, 343 (2022).

[62] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[63] There are two instances where a single exception occurs, but these exceptions can be attributed to incomplete optimization. See the Supplemental Material [57] for details.

[64] S. T. Flammia, Averaged circuit eigenvalue sampling, in *17th Conference on the Theory of Quantum Computation, Communication and Cryptography* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik 2022).

[65] https://github.com/kattemolle/HQAOA.

[66] F. G. S. L. Brandão, M. Broughton, E. Farhi, S. Gutmann, and H. Neven, For fixed control parameters the quantum approximate optimization algorithm's objective function value concentrates for typical instances, arXiv:1812.04170.