

Study of noise in virtual distillation circuits for quantum error mitigation

Pontus Vikstål¹, Giulia Ferrini¹, and Shruti Puri^{2,3}

¹Wallenberg Centre for Quantum Technology, Department of Microtechnology and Nanoscience, Chalmers University of Technology, 412 96 Gothenburg, Sweden

²Department of Applied Physics, Yale University, New Haven, Connecticut 06511, USA

³Yale Quantum Institute, Yale University, New Haven, Connecticut 06511, USA

Virtual distillation has been proposed as an error mitigation protocol for estimating the expectation values of observables in quantum algorithms. It proceeds by creating a cyclic permutation of M noisy copies of a quantum state using a sequence of controlled-swap gates. If the noise does not shift the dominant eigenvector of the density operator away from the ideal state, then the error in expectation-value estimation can be exponentially reduced with M . In practice, subsequent error mitigation techniques are required to suppress the effect of noise in the cyclic permutation circuit itself, leading to increased experimental complexity. Here, we perform a careful analysis of the effect of uncorrelated, identical noise in the cyclic permutation circuit and find that the estimation of expectation value of observables are robust against dephasing noise. We support the analytical result with numerical simulations and find that 67% of errors are reduced for $M = 2$, with physical dephasing error probabilities as high as 10%. Our results imply that a broad class of quantum algorithms can be implemented with higher accuracy in the near-term with qubit platforms where non-dephasing errors are suppressed, such as superconducting bosonic qubits and Rydberg atoms.

1 Introduction

Fault-tolerant quantum error correction is necessary for scalable quantum computation [1], however the associated hardware-performance requirements and resource overheads are hard to meet with the noisy intermediate-scale quantum processors available today. Consequently, for near-term applications alternative techniques to mitigate the effect of noise have been developed. Some of these techniques are based on scaling noise [2, 3, 4, 5] or learning about the effect of noise to predict the noise-free behavior of the

quantum protocol [6, 7], while others exploit the symmetry properties of the noise-free quantum circuit to flag errors [8, 9, 10, 11, 12]. Algorithm- and noise-specific error mitigation techniques have also been proposed [13, 14].

Recently an error mitigation scheme known as virtual distillation, or error suppression by derangement, has been shown to achieve an exponential suppression of errors in the estimation of the expectation value of an observable [15, 16, 17]. The key idea behind this protocol is to compute the expectation value of an observable by performing measurements on a cyclic-permutation of M copies of a noisy quantum state. If the effect of noise is to mix the ideal noise-free state with orthogonal error states, then symmetries of the cyclic-permutation state suppress the contribution to the expectation value from the error states exponentially in M .

The most straightforward approach to virtual distillation is to prepare the cyclic-permutation state using an auxiliary qubit and controlled-SWAP (CSWAP) gates. In practice, this circuit will be prone to errors, limiting the accuracy of expectation-value estimation without resorting to further noise mitigation techniques, like zero-noise extrapolation [2, 3, 16]. However, zero-noise extrapolation not only adds to the sampling cost, but also considerably increases the circuit complexity as it requires the ability to scale the noise strength in the quantum circuit either by scaling gate times or by adding more gates into the circuit [4, 5, 18, 19, 20]. Thus, in this paper we further investigate the effect of noise in the virtual distillation circuit and determine analytically conditions under which its faults may be less detrimental, obviating the need for additional error mitigation. We corroborate our findings with numerical simulations of the Quantum Approximate Optimization Algorithm (QAOA). Noise in virtual distillation circuits was previously considered numerically in the context of Heisenberg quench [17] as well as for the variational quantum eigensolver [21], displaying robustness of the error mitigation procedure.

We consider three commonly studied types of noise: depolarizing, dephasing, and amplitude damp-

Pontus Vikstål: e-mail: vikstal@chalmers.se

ing noise and find that the mitigated expectation value with virtual distillation is robust against dephasing noise for an arbitrary even number of copies M . We support our analysis with numerical simulation of QAOA for solving a MaxCut problem of partitioning the set of vertices in a given graph into two subsets such that the number of edges shared between the two partitions is maximized, for the case of two copies, $M = 2$. QAOA is implemented by preparing a variational quantum state using a short-depth quantum circuit and estimating its energy, i.e. the expectation value of the Ising-Hamiltonian associated with the MaxCut problem. The parameters in the circuit are varied until a minimum in the energy landscape is found. In order to overcome the adverse effects of noise in finding a state that minimizes the energy, we combine the QAOA protocol with virtual distillation. We find that the error in estimating the energy with virtual-distillation with $M = 2$ is reduced by 67% when the underlying source of noise is single-qubit pure dephasing errors at rate of 10%, compared to 20% when the underlying source of noise is single-qubit depolarizing errors at the same rate. Additionally, we found that amplitude damping errors was detrimental to the virtual distillation circuit, resulting in no error mitigation. Our findings imply that virtual distillation in a system in which non-dephasing errors are suppressed compared to dephasing errors is successful at reducing errors in expectation-value estimation of observables diagonal in the computational basis without additional error mitigation schemes. It is known that such an error channel is relevant for Kerr-cat qubits in superconducting microwave circuits [22, 23] and Rydberg atomic qubits [24] not only when the qubits are idle but also during implementation of Toffoli and controlled-not gates. These two gates can be combined to implement a CSWAP [25] and thus it is possible to realize robust virtual distillation in these platforms.

This paper is organized as follows: In Section 2 we review the virtual distillation protocol. We analyze the effect of noise in the virtual distillation circuit on the estimated expectation value in Section 3. We support our analysis with numerical simulations in Section 4 and finally, we give our concluding remarks in Section 5.

2 Virtual Distillation

We begin this section by establishing the notation used throughout this paper, which is based on Ref. [17]. A boldfaced superscript, for example O^i , will be used to indicate that the operator O acts on the i^{th} subsystem. We use superscript with parentheses to indicate an operator acting on multiple subsystems. For instance, $S^{(M)}$ indicates that S acts on M subsystems.

Consider the output density operator, ρ , of an N -

qubit noisy quantum circuit with the spectral decomposition

$$\rho = \sum_{k=1}^d \lambda_k |\psi_k\rangle\langle\psi_k|. \quad (1)$$

Here $d = 2^N$ and λ_k is the probability that the system is found in the state $|\psi_k\rangle$ when measuring in the eigenbasis of ρ . We assume, for convenience, that the probabilities λ_k are listed in descending order $\lambda_1 > \lambda_2 \dots > \lambda_d$. In the virtual distillation protocol, raising ρ to the power of M and normalizing it results in a density operator that approaches the dominant eigenvector $|\psi_1\rangle\langle\psi_1|$ exponentially fast with M , i.e.

$$\tilde{\rho} = \frac{\rho^M}{\text{Tr}(\rho^M)} = \frac{\sum_{k=1}^d \lambda_k^M |\psi_k\rangle\langle\psi_k|}{\sum_{k=1}^d \lambda_k^M}. \quad (2)$$

In virtual distillation, the expectation value of an observable O is estimated with respect to the exponentiated density matrix $\tilde{\rho}$,

$$\langle O \rangle_{\text{mitigated}} := \text{Tr}(O\tilde{\rho}) = \frac{\text{Tr}(O\rho^M)}{\text{Tr}(\rho^M)}. \quad (3)$$

When $|\psi_1\rangle$ corresponds to the output of the ideal (noise free) quantum circuit, then $\langle O \rangle_{\text{mitigated}}$ approaches the ideal expectation value exponentially fast with M . This condition is satisfied when noise in the quantum circuit maps the ideal states to states that are orthogonal to it, otherwise the dominant eigenvector will drift and limit the error suppression efficiency [16, 17]. In general, for a multi-qubit state, single-qubit errors can cause drift of the dominant eigenvector. However, in real-world applications, this drift is expected to be small, as also validated by the numerical simulations in this paper. Furthermore, the severity of this drift, or coherent mismatch, is exponentially smaller than the incoherent decay of fidelity [26].

Note that, in virtual distillation, $\langle O \rangle_{\text{mitigated}}$ is calculated without explicitly preparing the state $\tilde{\rho}$, hence the name “virtual”. Instead, virtual distillation uses M copies of the state ρ together with collective measurements that only allow symmetric states of the form $|\psi\rangle \otimes |\psi\rangle \dots |\psi\rangle$ to contribute to the expectation value of O . More specifically, in Ref. [17] it was shown that Eq. (3) is equivalent to

$$\langle O \rangle_{\text{mitigated}} := \frac{\text{Tr}(O^{(M)} S^{(M)} \rho^{\otimes M})}{\text{Tr}(S^{(M)} \rho^{\otimes M})}, \quad (4)$$

where $O^{(M)}$ is the symmetrized version of the operator O ,

$$O^{(M)} = \frac{1}{M} \sum_{i=1}^M O^i, \quad (5)$$

and $S^{(M)}$ is the cyclic shift operator that act on all M subsystems. Its effect is only to let symmetric

states of $\rho^{\otimes M}$ to contribute to the expectation value of Eq. (4),

$$S^{(M)} |\psi_1\rangle \otimes |\psi_2\rangle \dots |\psi_M\rangle = |\psi_2\rangle \otimes |\psi_3\rangle \dots |\psi_1\rangle. \quad (6)$$

To measure the observable $S^{(M)}$ in Eq. (4) virtual distillation uses a procedure similar to the Hadamard test [27]. The procedure begins by preparing M collective copies of the state ρ together with an auxiliary qubit in the state $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Next, a sequence of CSWAP gates applies $S^{(M)}$ to the M copies of ρ conditioned on the auxiliary qubit being in state $|1\rangle$. Finally the auxiliary qubit is measured in the X -basis and its expectation value equals to $\text{Tr}(S^{(M)}\rho^{\otimes M})$, i.e. the denominator of Eq. (4). Since $O^{(M)}$ commutes with $S^{(M)}$, these two operators can be simultaneously diagonalized, allowing them to be measured at the same time. By also measuring $O^{(M)}$ on the M subsystems ρ , the measurement outcome can be used together with the measurement outcome from the auxiliary qubit to estimate the numerator $\text{Tr}(O^{(M)}S^{(M)}\rho^{\otimes M})$.

3 Noise in virtual distillation circuits

We model a noisy gate in the virtual distillation circuit as an ideal gate followed by independent and identical single-qubit errors acting on each qubit participating in the gate. We examine three types of single-qubit noise channels. The first one is the depolarizing channel which describes a process where information is completely lost with some probability ϵ , and is given by [28]

$$\Lambda_{\text{dep}}(\rho) = (1 - \epsilon)\rho + \frac{\epsilon}{3}(X\rho X + Y\rho Y + Z\rho Z), \quad (7)$$

where $\{X, Y, Z\}$ are the Pauli operators and ϵ is the error probability. The second one is the pure-dephasing channel which is a biased noise channel¹ and describes loss of phase information with a probability ϵ ,

$$\Lambda_Z(\rho) = (1 - \epsilon)\rho + \epsilon Z\rho Z. \quad (8)$$

The third and final channel that we consider is the amplitude damping channel which is characterized by energy dissipation to the ground state over time. Although no analytical expression of the mitigated expectation value for the amplitude damping channel is derived in this work, it is defined as follows:

$$\Lambda_{\text{amp}}(\rho) = K_1\rho K_1^\dagger + K_2\rho K_2^\dagger. \quad (9)$$

where the Kraus operators K_1 and K_2 are given by

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}. \quad (10)$$

¹Of course we could have chosen an error channel with biased X - or Y -noise but we can always redefine the computational basis states on Bloch sphere and call all of these Z -biased noise.

with

$$\gamma \equiv 4(\sqrt{1-\epsilon} + \epsilon - 1). \quad (11)$$

We use these definitions of the error channels because their average channel fidelities are the same for a given ϵ , allowing for a consistent comparison across the different noise models.

In the next section we will present analytical results on how the depolarizing and dephasing noise channels affect the mitigated expectation value of virtual distillation as well as their associated variances. For the amplitude damping channel, instead, corresponding analytical expressions could not be obtained, and numerical results on that channel will be presented in the later Section 4.

3.1 Noisy mitigated expectation values

In this section we provide the main results of this paper. We derive an expression for the noisy mitigated expectation value for even number of copies. For any number of copies M , the cyclic shift operator $S^{(M)}$ factorizes into a tensor product of $MN/2$ number of SWAPs, and its controlled version factorizes into a product of $MN/2$ CSWAP-gates. For the virtual distillation circuit, we assume that a single-qubit noise channel is applied after each gate to the qubits involved, see FIG. 1. For even number of copies M , only one swap per subsystem is required, as for example shown for the case of $M = 4$ in FIG. 1b. As a consequence, for the case of even number of copies we find the following analytical expression for the mitigated expectation value:

$$\langle O \rangle_{\text{mitigated}}^{\Lambda} = \frac{\text{Tr}(\bar{\Lambda}(O)\rho^M)}{\text{Tr}(\rho^M)}, \quad (12)$$

where $\bar{\Lambda} = \Lambda \otimes \dots \otimes \Lambda$ is a tensor product of N single-qubit error channels and $\Lambda \in \{\Lambda_{\text{dep}}, \Lambda_Z\}$. The details of the calculations are provided in Appendix A. For the case of an odd number of copies, the calculation is more involved. Consider for instance the case of three copies. In this case, one of the copies needs to be swapped twice, making the mathematical derivation of the mitigated expectation value significantly more difficult. From Eq. (12) we see that the influence of errors on the mitigated expectation value will depend on the observable O . Since a general observable on N qubits can be expressed as a sum of N -qubit Pauli strings from the set $\{I, X, Y, Z\}^{\otimes N}$, it is sufficient to consider $O \in \{I, X, Y, Z\}^{\otimes N}$. In this case we find that the mitigated expectation value for the two types of noise are given by

$$\langle O \rangle_{\text{mitigated}}^{\Lambda_{\text{dep}}} = \left(1 - \frac{4}{3}\epsilon\right)^k \frac{\text{Tr}(O\rho^M)}{\text{Tr}(\rho^M)}, \quad (13)$$

$$\langle O \rangle_{\text{mitigated}}^{\Lambda_Z} = (1 - 2\epsilon)^{k'} \frac{\text{Tr}(O\rho^M)}{\text{Tr}(\rho^M)}, \quad (14)$$

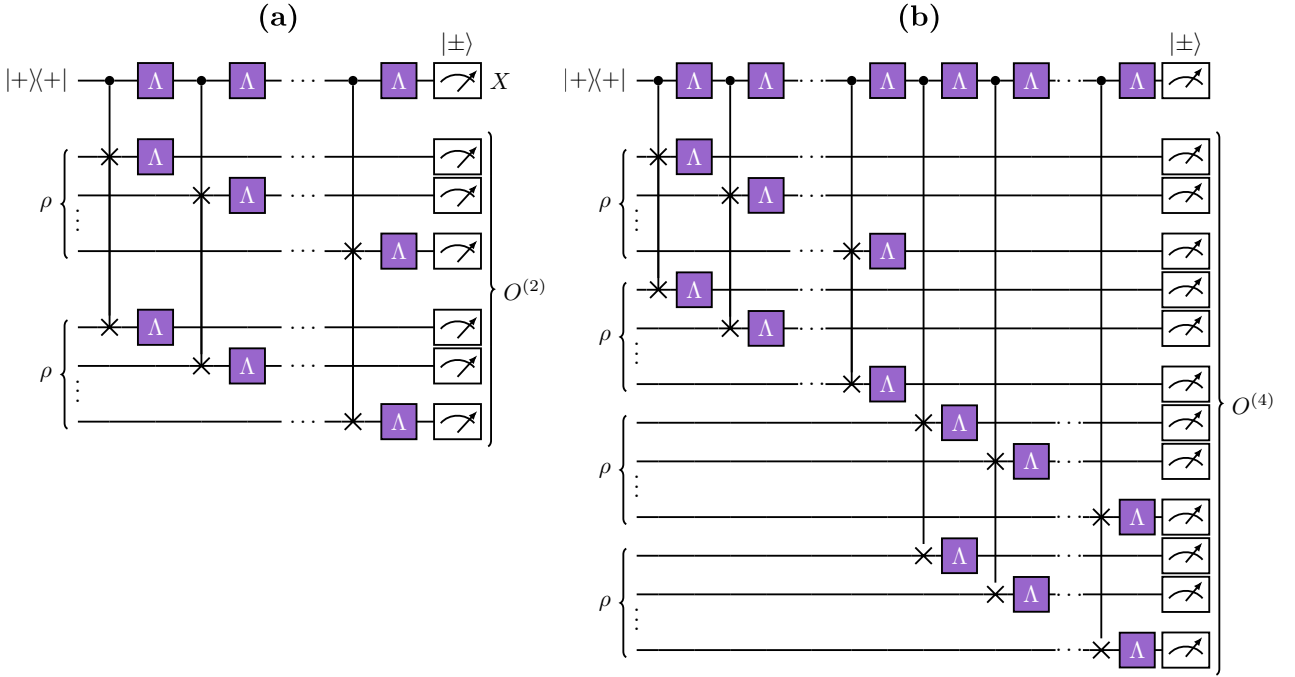


Figure 1: The virtual distillation circuit for **(a)** $M = 2$ and **(b)** $M = 4$, with single-qubit errors Λ .

where k is the number of $\{X, Y, Z\}$ Pauli matrices in the tensor product of O , and k' is the number of $\{X, Y\}$ Pauli matrices in the tensor product of O . Thus we see that errors in the virtual distillation circuit only attenuate the expectation values. We will study the sampling cost of this attenuation in the next section. The attenuation can, in principle, be overcome by error mitigation techniques like polynomial extrapolation. Importantly, we note that the mitigated expectation value of an observable O that is only a tensor product of Pauli Z -operators, so that $k' = 0$, is completely immune to pure dephasing in the distillation circuit. This is typically the case for variational algorithms for combinatorial optimization [29, 30, 31] or electronic structure calculation of molecular Hamiltonians [32]. Moreover by performing a local Clifford transformations on the state ρ before sending it into the virtual distillation circuit, it is possible to measure any Pauli observable without attenuating the mitigated expectation value.

3.2 Variance of the estimator

We now turn our attention to investigating the sample variance in the estimation of $\text{Tr}(O^{(M)} S^{(M)} \rho^{\otimes M}) / \text{Tr}(S^{(M)} \rho^{\otimes M})$. Here we limit ourselves to the case of $M = 2$, because the case of general number of copies involves calculating expectation values of the

sort $\langle (O^{(M)} X_{\text{aux}})^2 \rangle$ (where X_{aux} is the Pauli- X on the auxiliary qubit), which becomes quickly difficult for $M > 2$.

There exists no closed expression of the sample variance of the quotient between two random variables, but an approximated one can be obtained by Taylor-expanding the variance around the mean. If we let $\bar{x}$ denote the sample mean of $\text{Tr}(O^{(2)} S^{(2)} \rho^{\otimes 2})$, and $\bar{y}$ the sample mean of $\text{Tr}(S^{(2)} \rho^{\otimes 2})$, then the following unbiased estimator can be constructed

$$\tilde{\theta} := \frac{\bar{x}}{\bar{y}}. \quad (15)$$

Given sufficiently many samples R , the variance of this estimator, for the noiseless case of FIG. 1a, can be approximated as [17]

$$\text{Var}(\text{estim.}) \approx \frac{1}{R} \left(\frac{\text{Tr}(O^2 \rho)}{2 \text{Tr}(\rho^2)^2} + \frac{\text{Tr}(O \rho)^2}{2 \text{Tr}(\rho^2)^2} + \frac{\text{Tr}(O \rho^2)^2}{\text{Tr}(\rho^2)^4} - 2 \frac{\text{Tr}(O \rho^2)}{\text{Tr}(\rho^2)^3} \text{Tr}(O \rho) \right). \quad (16)$$

Our next step is to find how this variance changes when there is noise in the virtual distillation circuit. Starting with dephasing errors, we find in Appendix B that the variance of the estimator for a Pauli-string is

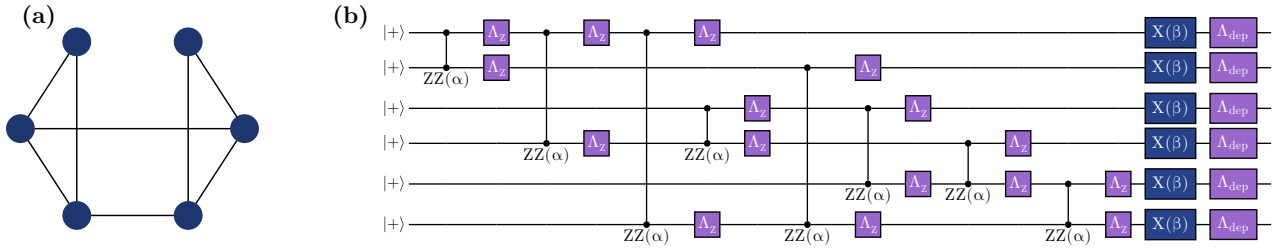


Figure 2: **(a)** A graph illustrating one instance of the MaxCut problem used in our study. Nodes represent qubits and edges denote interactions between them. **(b)** The corresponding QAOA circuit for $p = 1$ with dephasing noise for the graph in (a). Here $ZZ(\alpha)$ on qubit j and k is $e^{-i\alpha Z_j Z_k/2}$, and $X(\beta)$ on qubit j is $e^{-i\beta X_j}$. Λ_Z is a single qubit dephasing-channel and Λ_{dep} is a single-qubit depolarizing channel to account for the fact that Z -errors do not commute with the X -rotations. The total gate count in the figure is 6 single-qubit gates and 8 two-qubit gates, excluding the error-channels.

$$\text{Var}_{\Lambda_Z}(\text{estim.}) = \frac{1}{R(1-2\epsilon)^{2N}} \left(\frac{\text{Tr}(\bar{\Lambda}_Z(O^2)\rho)}{2\text{Tr}(\rho^2)^2} + \frac{\text{Tr}(\bar{\Lambda}_Z(O)\rho)^2}{2\text{Tr}(\rho^2)^2} + \frac{\text{Tr}(\bar{\Lambda}_Z(O)\rho^2)^2}{\text{Tr}(\rho^2)^4} - 2 \frac{\text{Tr}(\bar{\Lambda}_Z(O)\rho^2)}{\text{Tr}(\rho^2)^3} \text{Tr}(\bar{\Lambda}_Z(O)\rho) \right), \quad (17)$$

which is the same as Eq. (16) but scaled by a factor $(1-2\epsilon)^{-2N}$ and with O replaced by $\bar{\Lambda}_Z(O)$. In the case of O being a tensor product of Z -operators, $\bar{\Lambda}_Z(O) = O$, the expression reduces to

$$\text{Var}_{\Lambda_Z}(\text{estim.}) = \frac{1}{(1-2\epsilon)^{2N}} \text{Var}(\text{estim.}). \quad (18)$$

This equation shows that $(1-2\epsilon)^{2N}$ extra circuit repetitions are required for reaching the same level of precision compared to noise free virtual distillation. However, when $\epsilon \ll 1$ the denominator can be expanded as $(1-2\epsilon)^{-2N} \approx 1 + 4N\epsilon + \mathcal{O}(\epsilon^2)$, which shows that $4N\epsilon$ extra circuit repetitions are required for small error probabilities. In the presence of depolarizing errors no simple expression for the variance of the estimator is found, but we note that for a global depolarizing channel acting after each CSWAP gate a similar expression to Eq. (18) was found in Ref. [33].

4 Numerical Results

To corroborate our analytical results we will now perform numerical experiments by simulating the quantum approximate optimization algorithm (QAOA), solving 6-qubit MaxCut problems on 30 randomly generated Erdős-Rényi graphs [34] with 6-vertices and edge probability of 0.5 (see FIG. 2a).

The MaxCut problem is defined by a graph $G = (V, E)$, where V is the set of vertices and E is the set of edges. The objective of MaxCut is to partition the set of vertices into two subsets, such that the number of edges from one partition to the other is maximum. The problem can be reformulated as finding

the ground state of a Hamiltonian

$$H_{\text{MaxCut}} = -\frac{1}{2} \sum_{i,j \in E} (I - Z_i Z_j), \quad (19)$$

where Z_i and Z_j are Pauli Z matrices.

We simulate both QAOA and the virtual distillation circuits in the presence of either single-qubit depolarizing, dephasing or amplitude damping errors. Moreover, in order to also make a comparison that is independent of the state input to the three noisy virtual distillation circuits, we use the same 30 graphs to create a mixed state involving the state corresponding to the maximum cut and a thermal state for each graph. The mixed state is then used as input to the virtual distillation circuit, and we benchmark its performance in the presence of either single-qubit depolarizing, dephasing or amplitude damping errors. Regarding the complexity of our simulation, it should be noted that simulating the virtual distillation circuit for the two copy 6-qubit systems translates into a $2^{13} \times 2^{13}$ density matrix simulation, which is equivalent to a 26 qubit pure state simulation.

4.1 Virtual distillation applied to variational states

To approximate the ground state of H_{MaxCut} with QAOA, the variational state

$$\rho(\alpha, \beta) = U(\alpha, \beta)(|+\rangle\langle+|)^{\otimes N} U^\dagger(\alpha, \beta) \quad (20)$$

is prepared, where

$$U(\alpha, \beta) = \prod_{j=1}^p e^{-i\beta_j H_M} e^{-i\alpha_j H_{\text{MaxCut}}} \quad (21)$$

is a unitary operation, $H_M = \sum_i X_i$ is a sum of Pauli X matrices, $|+\rangle^{\otimes N}$ is a uniform superposition of all computational basis states, and α_j, β_j are $2p$ variational parameters. The variational parameters are optimized with respect to the expectation value of the MaxCut Hamiltonian

$$C(\alpha, \beta) := \text{Tr}(H_{\text{MaxCut}} \rho(\alpha, \beta)), \quad (22)$$

such that its value is minimized

$$(\alpha_{\text{opt}}, \beta_{\text{opt}}) := \arg \min_{\alpha, \beta} C(\alpha, \beta). \quad (23)$$

We simulate a noisy QAOA circuit for $p = 1$ where $e^{-i\alpha H_{\text{MaxCut}}}$ is implemented as a product of ZZ-rotations, $\prod_{j,k \in E} e^{-i\alpha Z_j Z_k / 2}$, and the mixer $e^{-i\beta H_M}$ is implemented as single qubit X-rotations, $\prod_j e^{-i\beta X_j}$. For the dephasing channel we have implemented the single-qubit error channel Λ_{dep} after the mixer gate to take into consideration the fact that dephasing errors do not commute with the X-rotation gate, resulting in an effect on the noise channel that will depolarize it and make it less noise biased, see FIG. 2b. In practical qubit platforms, two-qubit gates are generally more noisy than single-qubit gates [35]. Thus, we reduce the error probability ϵ by a factor 10 for the single-qubit gates. When simulating the QAOA with the amplitude damping channel Λ_{amp} , we use it for both the one and two-qubit gates with 10% less error probability for the single qubit gates. Finally, the same single-qubit error channel Λ that is used in the QAOA circuit is also used in the virtual distillation circuit. We label the noisy expectation value for error channel obtained using QAOA without virtual distillation as $C^\Lambda(\alpha, \beta)$ and with virtual distillation as $C_{\text{mitigated}}^\Lambda(\alpha, \beta)$.

We find the optimal variational parameters for both $C^\Lambda(\alpha, \beta)$ and $C_{\text{mitigated}}^\Lambda(\alpha, \beta)$ for 21 different error probabilities ϵ that are equally spaced between 0 and 0.1. To do this we start by optimizing the energy expectation value for no errors using brute-force optimization on a 100×100 grid with $\alpha \in [0, \pi]$ and $\beta \in [0, \pi/2]$ together with an optimization/polishing function that uses the optimal grid point as an initial guess. The optimal parameters found for the noiseless QAOA circuit are then used as initial guess to the optimization function for the first noise iteration, and the initial guess is iteratively updated for each increasing noise level based on the optimal parameters found for the previous noise level. With this strategy, we aim to reduce the search space and computational time required for optimization at each subsequent noise level. Previous studies have indeed suggested that optimal

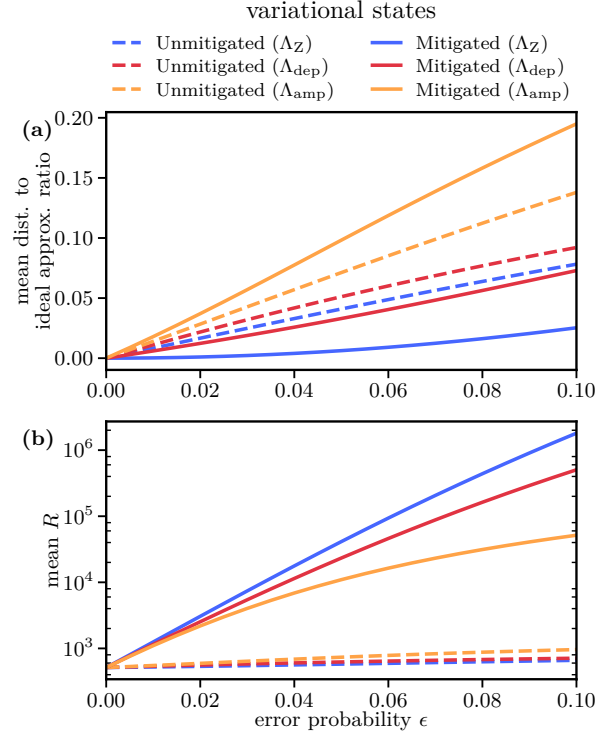


Figure 3: **(a)** The distance from the ideal approximation ratio with respect to the error probability ϵ averaged over all the instances. **(b)** The minimum number of samples R required for the variance of the estimator to be $\leq 10^{-3}$ averaged over all the instances. Solid (dashed) lines are obtained with (without) virtual distillation. Blue lines are with respect to dephasing errors, red lines are with respect to depolarizing errors, and yellow lines with respect to amplitude damping.

parameters for QAOA circuits tend to remain relatively stable in the presence of moderate noise [36, 37], providing another rationale for our iterative optimization approach.

After the optimization we compute the approximation ratio by dividing the expected cuts $C(\alpha_{\text{opt}}, \beta_{\text{opt}})$, and $C_{\text{mitigated}}^\Lambda(\alpha_{\text{opt}}, \beta_{\text{opt}})$ by the maximum cut C_{max} , where both expected cuts are calculated explicitly. FIG. 3a presents the difference between the ideal (noiseless) approximation ratio and the noisy approximation ratio averaged over the 30 instances without virtual distillation (unmitigated) for QAOA with the three types of error-channels, and with virtual distillation (mitigated) for the three types of error-channels. It can clearly be seen that the best error reduction is obtained for dephasing-errors in both the QAOA circuit and the virtual distillation circuit. From the figure the error reduction at $\epsilon = 0.1$ with virtual distillation is 67% when the underlying source of noise is dephasing errors, which shows excellent error suppression given a high error probability ϵ . For depolarizing errors the error reduction is 20% with virtual distillation. For amplitude damping errors we see no improvement, and actually the virtual distillation proce-

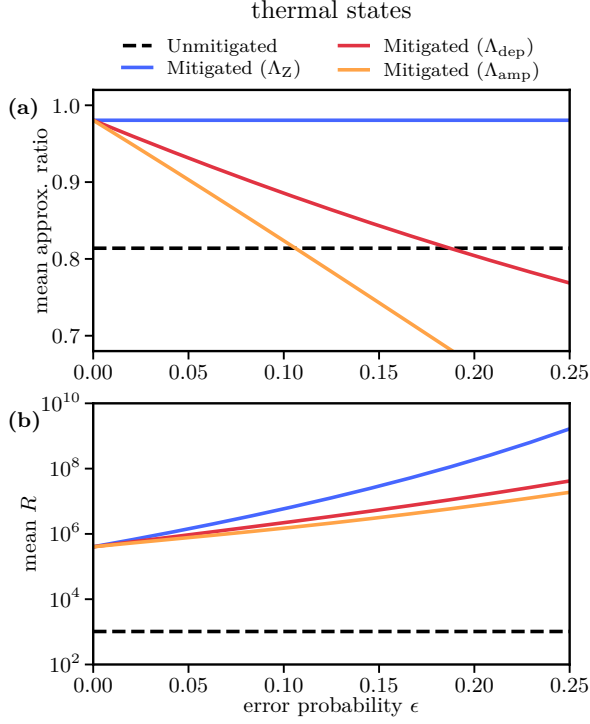


Figure 4: **(a)** The approximation ratio with respect to the error probability ϵ averaged over all the instances. **(b)** The minimum number of samples R required for the variance of the estimator to be $\leq 10^{-3}$ averaged over all the instances. Solid (dashed) lines are obtained with (without) virtual distillation. Blue lines are with respect to dephasing errors, red lines are with respect to depolarizing errors, and yellow lines with respect to amplitude damping.

ture degrades the results. This phenomenon presents an intriguing aspect of quantum error dynamics in the virtual distillation circuit that is not yet fully understood. Interestingly, we find that dephasing-errors in the QAOA circuit yield a moderately better mean approximation ratio compared to depolarizing-errors without virtual distillation, as seen by the blue dashed line being below the red dashed line. This indicates that QAOA by itself may also be more robust against dephasing errors.

It is known that single-qubit errors can lead to a coherent mismatch of the dominant eigenvector with respect to the ideal state [16, 17, 26]. This means that the fidelity between dominant eigenvector of ρ and the ideal noiseless state is not unity. In Appendix C we do a numerical analysis of the coherent mismatch caused by the errors in the QAOA circuit, and find that it is smaller for dephasing errors in the circuit.

In FIG. 3b we plot the minimum number of samples R required for the variance of the estimator defined in Eq. (15) to be $\leq 10^{-3}$, averaged over all the 30 instances. We also do this for the unmitigated variance

of the sample mean, given by

$$\text{Var}(\text{estim.}) = \frac{1}{2R} \left[\text{Tr}(\rho_{\Lambda} H_{\text{MaxCut}}^2) - \text{Tr}(\rho_{\Lambda} H_{\text{MaxCut}})^2 \right], \quad (24)$$

where R is the total number of samples and ρ_{Λ} is the output from the QAOA circuit with error channel $\Lambda \in \{\Lambda_{\text{dep}}, \Lambda_Z\}$. Since the virtual distillation circuit uses two copies of ρ_{Λ} as input but is only counted as one sample, therefore to provide an adjusted comparison that takes into account the additional resources used in virtual distillation, we scale the variance of the unmitigated sample mean by $1/2$. Indeed, for a pure state ρ , it can be shown that Eq. (16) is equal to Eq. (24) [17]. From the two solid lines in FIG. 3b, we see that the mean number of samples grow exponentially with virtual distillation for both types of errors. However, the difference between the number of samples needed for dephasing noise is only negligibly ($\ll$ factor of 10) larger than those needed for depolarizing noise.

4.2 Virtual distillation applied to thermal states

In the previous section, the density matrices that was given as input to the virtual distillation circuit were different as a result of the error-channel in the QAOA circuit. In this section, we make a noise agnostic comparison where the input state is the same to both virtual distillation circuits. The state that we have chosen is a statistical mixture between the state corresponding to the maximum-cut and a thermal state. The state corresponding the maximum-cut is given by the degenerate ground state $\rho_{\text{ideal}} = |\psi_{\text{GS}}\rangle\langle\psi_{\text{GS}}|$ of the MaxCut Hamiltonian. We then consider the thermal state

$$\rho_{\text{thermal}} = \frac{e^{-\eta H_{\text{MaxCut}}}}{Z(\eta)}, \quad (25)$$

where $Z(\eta) = \text{Tr}(e^{-\eta H_{\text{MaxCut}}})$ is the partition function, and η is a constant that is proportional to the inverse temperature. From this we create an equally mixed state between the two states ρ_{ideal} and ρ_{thermal} ,

$$\rho = \frac{1}{2}\rho_{\text{ideal}} + \frac{1}{2}\rho_{\text{thermal}}. \quad (26)$$

Using ρ as our input to the virtual distillation circuit we perform virtual distillation on it using the circuit in FIG. 1a with Λ as either Λ_Z , Λ_{dep} or Λ_{amp} . In the simulations we choose $\eta = 0.1$, and vary the noise-level ϵ between 0 and 0.25 and calculate the mitigated approximation ratio as well as the number of repetitions R required for the variance of the estimator to be $\leq 10^{-3}$, averaged over all the 30 instances. FIG. 4a shows the mean approximation ratio. While the mitigated approximation ratio stays constant for dephasing errors irrespective of the error probability

ϵ , the mitigated approximation ratio for both depolarizing errors and amplitude damping quickly decreases and even becomes lower than the unmitigated approximation ratio as seen by the solid red and yellow line crossing the black dashed line. This is because, while the MaxCut Hamiltonian commutes with the error operator Z for the pure dephasing channel, it does not commute with all the error operators for the depolarizing and amplitude damping one. FIG. 4b shows that the mean number of samples for the mitigated expectation values grows rapidly with the error probability. However, as before, the number of samples required when noise is pure dephasing is only slightly larger ($\lesssim$ factor of 10) than those required when noise is depolarizing. Note that the two solid lines do not intercept the dashed line at $\epsilon = 0$, which is because the state ρ in Eq. (26) is not a pure state.

5 Conclusion

In this work, we have studied the effects of depolarizing-, dephasing-errors and amplitude damping in virtual distillation circuits. We found that depolarizing errors in the circuit implementation of virtual distillation substantially degrade the mitigated expectation value. Additionally, it was observed that amplitude damping is extremely detrimental for virtual distillation resulting in no error mitigation when estimating the expectation value of the MaxCut Hamiltonian. For dephasing errors, we found that the quality of the mitigated expectation value does not degrade. This makes virtual distillation implemented in a system that strongly favors biased noise particular robust to errors. Moreover, trading biased noise for depolarizing errors does not significantly increase the sampling cost.

Our findings suggest that implementing the virtual distillation protocol in a system that is strongly biased towards dephasing noise thus avoids the need to rely on other error mitigation techniques in addition to virtual distillation for mitigating errors. Finally, we mention that bias-preserving CSWAP gates can be implemented in bosonic cat code systems [22, 38] and Rydberg atoms [24], which is crucial in order to not unbiased or depolarize the noise channel.

Acknowledgments

We thank Oliver Hahn, Timo Hillman, Shahnawaz Ahmed, and Robert Jonsson for fruitful discussions. This work is supported from the Knut and Alice Wallenberg Foundation through the Wallenberg Center for Quantum Technology (WACQT). G. F. acknowledges support from the Swedish Research Council (Vetenskapsrådet) Grant QuACVA. SP was supported by the Air Force Office of Scientific Research under award number FA9550-21-1-0209.

Code Availability

The code used for producing the results is made available in Ref. [39]. All circuit simulations are done using Cirq [40] and Numpy [41]. As optimization function for the QAOA, we used MINIMIZE implemented in Scipy [42] with the BFGS algorithm as the default optimizer. The random graphs were generated using NetworkX [43].

References

- [1] Daniel A Lidar and Todd A Brun. “Quantum error correction”. Cambridge University Press. (2013).
- [2] Ying Li and Simon C. Benjamin. “Efficient variational quantum simulator incorporating active error minimization”. *Phys. Rev. X* **7**, 021050 (2017).
- [3] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. “Error mitigation for short-depth quantum circuits”. *Phys. Rev. Lett.* **119**, 180509 (2017).
- [4] Suguru Endo, Simon C Benjamin, and Ying Li. “Practical quantum error mitigation for near-future applications”. *Phys. Rev. X* **8**, 031027 (2018).
- [5] Abhinav Kandala, Kristan Temme, Antonio D Córcoles, Antonio Mezzacapo, Jerry M Chow, and Jay M Gambetta. “Error mitigation extends the computational reach of a noisy quantum processor”. *Nature* **567**, 491–495 (2019).
- [6] Armands Strikis, Dayue Qin, Yanzhu Chen, Simon C Benjamin, and Ying Li. “Learning-based quantum error mitigation”. *PRX Quantum* **2**, 040330 (2021).
- [7] Piotr Czarnik, Andrew Arrasmith, Patrick J Coles, and Lukasz Cincio. “Error mitigation with clifford quantum-circuit data”. *Quantum* **5**, 592 (2021).
- [8] Xavi Bonet-Monroig, Ramiro Sagastizabal, M Singh, and T. E. O’Brien. “Low-cost error mitigation by symmetry verification”. *Phys. Rev. A* **98**, 062339 (2018).
- [9] Sam McArdle, Xiao Yuan, and Simon Benjamin. “Error-mitigated digital quantum simulation”. *Phys. Rev. Lett.* **122**, 180501 (2019).
- [10] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O’Brien, and L. DiCarlo. “Experimental error mitigation via symmetry verification in a variational quantum eigensolver”. *Phys. Rev. A* **100**, 010302(R) (2019).

- [11] Google AI Quantum and Collaborators, Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B Buckley, et al. “Hartree-fock on a superconducting qubit quantum computer”. *Science* **369**, 1084–1089 (2020).
- [12] William J Huggins, Jarrod R McClean, Nicholas C Rubin, Zhang Jiang, Nathan Wiebe, K Birgitta Whaley, and Ryan Babbush. “Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers”. *npj Quantum Inf* **7**, 1–9 (2021).
- [13] Thomas E O’Brien, Stefano Polla, Nicholas C Rubin, William J Huggins, Sam McArdle, Sergio Boixo, Jarrod R McClean, and Ryan Babbush. “Error mitigation via verified phase estimation”. *PRX Quantum* **2**, 020317 (2021).
- [14] Filip B Maciejewski, Zoltán Zimborás, and Michał Oszmaniec. “Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography”. *Quantum* **4**, 257 (2020).
- [15] Jordan Cotler, Soonwon Choi, Alexander Lukin, Hrant Gharibyan, Tarun Grover, M. Eric Tai, Matthew Rispoli, Robert Schittko, Philipp M. Preiss, Adam M. Kaufman, Markus Greiner, Hannes Pichler, and Patrick Hayden. “Quantum virtual cooling”. *Phys. Rev. X* **9**, 031013 (2019).
- [16] Bálint Koczor. “Exponential error suppression for near-term quantum devices”. *Phys. Rev. X* **11**, 031057 (2021).
- [17] William J. Huggins, Sam McArdle, Thomas E. O’Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. “Virtual distillation for quantum error mitigation”. *Phys. Rev. X* **11**, 041036 (2021).
- [18] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski. “Cloud quantum computing of an atomic nucleus”. *Phys. Rev. Lett.* **120**, 210501 (2018).
- [19] Matthew Otten and Stephen K. Gray. “Recovering noise-free quantum observables”. *Phys. Rev. A* **99**, 012338 (2019).
- [20] Tudor Giurgica-Tiron, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J. Zeng. “Digital zero noise extrapolation for quantum error mitigation”. *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* Pages 306–316 (2020).
- [21] Hamza Jnane, Brennan Undseth, Zhenyu Cai, Simon C. Benjamin, and Bálint Koczor. “Multicore quantum computing”. *Phys. Rev. Appl.* **18**, 044064 (2022).
- [22] Shruti Puri, Lucas St-Jean, Jonathan A. Gross, Alexander Grimm, Nicholas E. Frattini, Pavithran S. Iyer, Anirudh Krishna, Steven Touzard, Liang Jiang, Alexandre Blais, Steven T. Flammia, and S. M. Girvin. “Bias-preserving gates with stabilized cat qubits”. *Sci Adv* **6**, eaay5901 (2020).
- [23] Alexander Grimm, Nicholas E Frattini, Shruti Puri, Shantanu O Mundhada, Steven Touzard, Mazhar Mirrahimi, Steven M Girvin, Shyam Shankar, and Michel H Devoret. “Stabilization and operation of a kerr-cat qubit”. *Nature* **584**, 205–209 (2020).
- [24] Iris Cong, Harry Levine, Alexander Keesling, Dolev Bluvstein, Sheng-Tao Wang, and Mikhail D. Lukin. “Hardware-efficient, fault-tolerant quantum computation with rydberg atoms”. *Phys. Rev. X* **12**, 021049 (2022).
- [25] John A. Smolin and David P. DiVincenzo. “Five two-bit quantum gates are sufficient to implement the quantum fredkin gate”. *Phys. Rev. A* **53**, 2855 (1996).
- [26] Bálint Koczor. “The dominant eigenvector of a noisy quantum state”. *New J. Phys.* **23**, 123047 (2021).
- [27] Dorit Aharonov, Vaughan Jones, and Zeph Landau. “A polynomial quantum algorithm for approximating the jones polynomial”. *Algorithmica* **55**, 395–421 (2009).
- [28] Michael A. Nielsen and Isaac L. Chuang. “Quantum computation and quantum information: 10th anniversary edition”. *Cambridge University Press*. (2010).
- [29] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm” (2014). [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [30] Andrew Lucas. “Ising formulations of many NP problems”. *Front. Physics* **2**, 5 (2014).
- [31] Pontus Vikstål, Mattias Grönkvist, Marika Svensson, Martin Andersson, Göran Johansson, and Giulia Ferrini. “Applying the quantum approximate optimization algorithm to the tail-assignment problem”. *Phys. Rev. Applied* **14**, 034009 (2020).
- [32] Rongxin Xia, Teng Bian, and Sabre Kais. “Electronic structure calculations and the ising hamiltonian”. *J. Phys. Chem. B* **122**, 3384–3395 (2018).
- [33] Piotr Czarnik, Andrew Arrasmith, Lukasz Cincio, and Patrick J. Coles. “Qubit-efficient exponential suppression of errors” (2021). [arXiv:2102.06056](https://arxiv.org/abs/2102.06056).
- [34] P. Erdős and A. Rényi. “On random graphs. i.”. *Publ. Math. Debrecen* **6**, 290–297 (1959).

- [35] Swamit S. Tannu and Moinuddin K. Qureshi. “Not all qubits are created equal”. In *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*. Pages 987–999. ACM (2019).
- [36] Cheng Xue, Zhao-Yun Chen, Yu-Chun Wu, and Guo-Ping Guo. “Effects of quantum noise on quantum approximate optimization algorithm” (2019). [arxiv:1909.02196](https://arxiv.org/abs/1909.02196).
- [37] Kunal Sharma, Sumeet Khatri, M Cerezo, and Patrick J Coles. “Noise resilience of variational quantum compiling”. *New J. Phys.* **22**, 043006 (2020).
- [38] Jérémie Guillaud and Mazyar Mirrahimi. “Repetition cat qubits for fault-tolerant quantum computation”. *Phys. Rev. X* **9**, 041053 (2019).
- [39] Pontus Wikstahl. “Wikstahl/error-mitigation: Version 2.0 (v2.0)”. Zenodo, <https://doi.org/10.5281/zenodo.7215577> (2022).
- [40] Cirq Developers. “Cirq (v0.12.0)”. Zenodo, <https://doi.org/10.5281/zenodo.5182845> (2021).
- [41] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. “Array programming with NumPy”. *Nature* **585**, 357–362 (2020).
- [42] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt aberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. *Nature Methods* **17**, 261–272 (2020).
- [43] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. “Exploring network structure, dynamics, and function using networkx”. Presented at the 7th Python in Science Conference, Pasadena, CA USA (2008). pp. 11-15.
- [44] G.M.P. van Kempen and L.J. van Vliet. “Mean and variance of ratio estimators used in fluorescence ratio imaging”. *Cytometry* **39**, 300–305 (2000).

A Error mitigated expectation value

In this section we derive an analytical expression of how the mitigated expectation value is affected by single-qubit depolarizing and dephasing errors that occur during the virtual distillation circuit. To simplify the derivation it is useful to use the following convention for the single-qubit error channels:

$$\Lambda_{\text{dep}}(\rho) = \left(1 - \frac{3}{4}\epsilon\right)\rho + \frac{3}{4}\epsilon(X\rho X + Y\rho Y + Z\rho Z), \quad (27)$$

$$\Lambda_Z(\rho) = \left(1 - \frac{\epsilon}{2}\right)\rho + \frac{\epsilon}{2}Z\rho Z, \quad (28)$$

which amounts to substituting $\epsilon \rightarrow 4\epsilon/3$ in the end to get back to Eq. (7) for the depolarizing channel, and $\epsilon \rightarrow 2\epsilon$ to get back to Eq. (8) for the dephasing channel.

We start by deriving the noisy mitigated expectation value for two-copies virtual distillation and then generalize the results to an even number of copies. The output from the virtual distillation circuit in FIG. 1a can be written as

$$\rho_{\text{out}} = \underbrace{\tilde{\Lambda}_N \circ \mathcal{U}_N \circ \tilde{\Lambda}_{N-1} \circ \mathcal{U}_{N-1} \dots \tilde{\Lambda}_1 \circ \mathcal{U}_1}_{\Lambda_{\text{tot}}} \rho_{\text{in}} = \Lambda_{\text{tot}}(\rho_{\text{in}}), \quad (29)$$

where ρ_{in} is the input state

$$\rho_{\text{in}} := |+\rangle\langle+| \otimes \rho \otimes \rho, \quad (30)$$

$\mathcal{U}_i(\rho_{\text{in}}) = \text{CSWAP}_i \rho_{\text{in}} \text{CSWAP}_i^\dagger$ is the controlled-SWAP gate that swaps the i th qubit of the two subsystems $\rho^{\otimes 2}$ conditioned on the auxiliary qubit being in the state $|1\rangle$, and is given by

$$\text{CSWAP}_i = |0\rangle\langle 0| \otimes I^{\otimes N} \otimes I^{\otimes N} + |1\rangle\langle 1| \otimes \text{SWAP}_i, \quad (31)$$

where I is a 2×2 identity matrix; $\tilde{\Lambda}_i$ is the product of the single-qubit error channels that are applied to the qubits involved in the gate $\mathcal{U}_i$,

$$\tilde{\Lambda}_i = \Lambda \otimes \Lambda_i \otimes \Lambda_i = \Lambda \otimes \Lambda_i^{\otimes 2}, \quad (32)$$

with $\Lambda \in \{\Lambda_{\text{dep}}, \Lambda_Z\}$ being a single-qubit error channel, and the subscript i of Λ_i indicates that Λ acts on the i th qubit of subsystem ρ . Using the cyclic permutation of the trace the expectation value of the numerator in Eq. (4) for $M = 2$ can be expressed as

$$\text{Tr}\left(X_{\text{aux}} O^{(2)} \Lambda_{\text{tot}}(\rho_{\text{in}})\right) = \text{Tr}\left(\Lambda_{\text{tot}}^\dagger(X_{\text{aux}} O^{(2)}) \rho_{\text{in}}\right), \quad (33)$$

where

$$\Lambda_{\text{tot}}^\dagger := \mathcal{U}_1^\dagger \circ \tilde{\Lambda}_1^\dagger \circ \mathcal{U}_2^\dagger \circ \tilde{\Lambda}_2^\dagger \dots \mathcal{U}_N^\dagger \circ \tilde{\Lambda}_N^\dagger = \circ_{k=1}^N (\mathcal{U}_k^\dagger \circ \tilde{\Lambda}_k^\dagger) \quad (34)$$

is the adjoint of Λ_{tot} , X_{aux} is the Pauli X operator on the auxiliary qubit, and $O^{(2)}$ acts on the two subsystems ρ . It can be noted that Eq. (33) is reminiscent of the Heisenberg picture in which the operators are evolving instead of the quantum states. We can furthermore drop $\dagger$ from $\mathcal{U}_k$ and Λ_k since they are both Hermitian quantum maps.

Beginning with the action of $\tilde{\Lambda}_N$ in Eq. (34) on $X_{\text{aux}} O^{(2)}$, the result is

$$\tilde{\Lambda}_N(X_{\text{aux}} O^{(2)}) = \Lambda(X) \otimes \Lambda_N^{\otimes 2}(O^{(2)}) = (1 - \epsilon)X \otimes \Lambda_N^{\otimes 2}(O^{(2)}), \quad (35)$$

where we have used the fact that

$$\Lambda(X) = \Lambda(|0\rangle\langle 1|) + \Lambda(|1\rangle\langle 0|) = (1 - \epsilon)|0\rangle\langle 1| + (1 - \epsilon)|1\rangle\langle 0| = (1 - \epsilon)X, \quad (36)$$

for $\Lambda \in \{\Lambda_{\text{dep}}, \Lambda_Z\}$. Next, Eq. (35) is followed by the action of $\mathcal{U}_N$ which gives

$$(\mathcal{U}_N \circ \tilde{\Lambda}_N)(X_{\text{aux}} O^{(2)}) = (1 - \epsilon) \left[|1\rangle\langle 0| \otimes \text{SWAP}_N \Lambda_N^{\otimes 2}(O^{(2)}) + |0\rangle\langle 1| \otimes \Lambda_N^{\otimes 2}(O^{(2)}) \text{SWAP}_N \right]. \quad (37)$$

Repeating this for the remaining terms yields

$$\begin{aligned} \Lambda_{\text{tot}}^\dagger(X_{\text{aux}} O^{(2)}) &= (1 - \epsilon)^N \left[|1\rangle\langle 0| \otimes \prod_{k=1}^N \text{SWAP}_k \tilde{\Lambda}^{\otimes 2}(O^{(2)}) + |0\rangle\langle 1| \otimes \tilde{\Lambda}^{\otimes 2}(O^{(2)}) \prod_{k=1}^N \text{SWAP}_k \right] \\ &= (1 - \epsilon)^N X \otimes S^{(2)} \tilde{\Lambda}^{\otimes 2}(O^{(2)}), \end{aligned} \quad (38)$$

where $\bar{\Lambda}^{\otimes 2} := \bar{\Lambda} \otimes \bar{\Lambda} = (\Lambda_1 \circ \dots \circ \Lambda_N) \otimes (\Lambda_1 \circ \dots \circ \Lambda_N) = \Lambda \otimes \dots \otimes \Lambda$ is a tensor product of $2N$ single-qubit error channels Λ , and where we have used the fact that $S^{(2)} = \prod_{k=1}^N \text{SWAP}_k$ and $[S^{(2)}, \bar{\Lambda}^{\otimes 2}(O^{(2)})] = 0$ in the last equality. The expectation value of the numerator is thus calculated to be

$$\text{Tr}\left(\Lambda_{\text{tot}}^\dagger(X_{\text{aux}}O^{(2)})\rho_{\text{in}}\right) = (1-\epsilon)^N \text{Tr}(X|+\rangle\langle+|) \text{Tr}\left(\bar{\Lambda}^{\otimes 2}(O^{(2)})S^{(2)}\rho^{\otimes 2}\right) = (1-\epsilon)^N \text{Tr}(\bar{\Lambda}(O)\rho^2). \quad (39)$$

To calculate the denominator of Eq. (4) for $M = 2$, we simply replace O with the identity $I^{\otimes N}$ for which $\bar{\Lambda}(I^{\otimes N}) = I^{\otimes N}$ because both the depolarizing- and dephasing-channel are unital maps. Hence the mitigated expectation value with either depolarizing or dephasing-errors is given by

$$\langle O \rangle_{\text{mitigated}}^\Lambda = \frac{\text{Tr}\left(\Lambda_{\text{tot}}^\dagger(X_{\text{aux}}O^{(2)})\rho_{\text{in}}\right)}{\text{Tr}\left(\Lambda_{\text{tot}}^\dagger(X_{\text{aux}})\rho_{\text{in}}\right)} = \frac{(1-\epsilon)^N \text{Tr}(\bar{\Lambda}(O)\rho^2)}{(1-\epsilon)^N \text{Tr}(\rho^2)} = \frac{\text{Tr}(\bar{\Lambda}(O)\rho^2)}{\text{Tr}(\rho^2)}. \quad (40)$$

We have reached the final expression which shows that single-qubit depolarizing or dephasing errors that occur during the circuit are equivalent to a tensor product of N -single qubit depolarizing or dephasing channels Λ acting on the observable O .

It is useful to consider what happens for the special case when O is a tensor product of Pauli-operators, e.g. $O = X \otimes Z \otimes I \otimes Y \dots$. In this case we have that $\Lambda_{\text{dep}}(\sigma) = (1-\epsilon)\sigma$, with $\sigma \in \{X, Y, Z\}$, for the depolarizing channel, and $\Lambda_Z(\sigma) = (1-\epsilon)\sigma$, with $\sigma \in \{X, Y\}$, for the dephasing channel. Hence the mitigated expectation value for the two channels can be written as

$$\langle O \rangle_{\text{mitigated}}^{\Lambda_{\text{dep}}} = (1-\epsilon)^k \frac{\text{Tr}(O\rho^2)}{\text{Tr}(\rho^2)}, \quad (41)$$

$$\langle O \rangle_{\text{mitigated}}^{\Lambda_Z} = (1-\epsilon)^{k'} \frac{\text{Tr}(O\rho^2)}{\text{Tr}(\rho^2)}, \quad (42)$$

where k is the number of $\{X, Y, Z\}$ Pauli-operators in the tensor product of O and k' is the number of $\{X, Y\}$ Pauli-operators in the tensor product of O . With the substitution of ϵ back to the definitions of Eq. (7) and Eq. (8) the mitigated expectation values becomes

$$\langle O \rangle_{\text{mitigated}}^{\Lambda_{\text{dep}}} = \left(1 - \frac{4}{3}\epsilon\right)^k \frac{\text{Tr}(O\rho^2)}{\text{Tr}(\rho^2)}, \quad (43)$$

$$\langle O \rangle_{\text{mitigated}}^{\Lambda_Z} = (1-2\epsilon)^{k'} \frac{\text{Tr}(O\rho^2)}{\text{Tr}(\rho^2)}. \quad (44)$$

These results can easily be generalized for arbitrary even M . This is because for even M one can execute controlled-swaps such that every subsystem ρ is only involved in one controlled-swap operations, which is not the case for odd M . For instance, with four copies of ρ , one can execute controlled-swaps between subsystems one and two, and three and four, respectively, as depicted in FIG. 1b. The importance is to ensure that no copy remains in its original position, thus achieving the necessary derangement. Consequently, the auxiliary qubit is engaged in $NM/2$ controlled-swaps. Therefore the expression for the mitigated expectation value for even M becomes

$$\langle O \rangle_{\text{mitigated}}^\Lambda = \frac{\text{Tr}\left(\Lambda_{\text{tot}}^\dagger(X_{\text{aux}}O^{(M)})\rho_{\text{in}}\right)}{\text{Tr}\left(\Lambda_{\text{tot}}^\dagger(X_{\text{aux}})\rho_{\text{in}}\right)} = \frac{(1-\epsilon)^{NM/2} \text{Tr}(\bar{\Lambda}(O)\rho^M)}{(1-\epsilon)^{NM/2} \text{Tr}(\rho^M)} = \frac{\text{Tr}(\bar{\Lambda}(O)\rho^M)}{\text{Tr}(\rho^M)}, \quad (45)$$

which is the same expression as Eq. (40), but for arbitrary even M .

B Variance of the estimator with dephasing errors

In this section, we derive the variance of the estimator in Eq. (15) with dephasing errors in the virtual distillation circuit for $M = 2$. While there exists no closed-form expression for the variance of the quotient between two random variables, an approximate expression can be obtained by doing a Taylor expansion around the mean of the two random variables [44]. By doing so, one finds that the variance of the estimator is

$$\text{Var}(\text{estim.}) \approx \frac{1}{R} \left(\frac{\text{Var}(X_{\text{aux}}O^{(2)})}{\langle X_{\text{aux}} \rangle^2} + \frac{\langle X_{\text{aux}}O^{(2)} \rangle^2}{\langle X_{\text{aux}} \rangle^4} \text{Var}(X_{\text{aux}}) - 2 \frac{\langle O^{(2)} X_{\text{aux}} \rangle}{\langle X_{\text{aux}} \rangle^3} \text{Cov}(O^{(2)} X_{\text{aux}}, X_{\text{aux}}) \right), \quad (46)$$

where R is the number of samples. The terms $\langle X_{\text{aux}} \rangle$ and $\langle O^{(2)} X_{\text{aux}} \rangle$ have been evaluated in [Appendix A](#), and we are thus left to evaluate the three following terms

$$\text{Var}(X_{\text{aux}}) = \langle (X_{\text{aux}})^2 \rangle - \langle X_{\text{aux}} \rangle^2 = 1 - (1 - 2\epsilon)^{2N} \text{Tr}(\rho^2)^2, \quad (47)$$

$$\begin{aligned} \text{Cov}(O^{(2)} X_{\text{aux}}, X_{\text{aux}}) &= \langle O^{(2)} X_{\text{aux}}^2 \rangle - \langle O^{(2)} X_{\text{aux}} \rangle \langle X_{\text{aux}} \rangle \\ &= \langle O^{(2)} \rangle - (1 - 2\epsilon)^{2N} \text{Tr}(\bar{\Lambda}_Z(O)\rho^2) \text{Tr}(\rho^2) \\ &= \text{Tr}(\bar{\Lambda}_Z(O)\rho) - (1 - 2\epsilon)^{2N} \text{Tr}(\bar{\Lambda}_Z(O)\rho^2) \text{Tr}(\rho^2). \end{aligned} \quad (48)$$

$$\begin{aligned} \text{Var}(O^{(2)} X_{\text{aux}}) &= \langle (O^{(2)} X_{\text{aux}})^2 \rangle - \langle O^{(2)} X_{\text{aux}} \rangle^2 \\ &= \langle (O^{(2)})^2 \rangle - (1 - 2\epsilon)^{2N} \text{Tr}(\bar{\Lambda}_Z(O)\rho^2)^2 \\ &= \frac{1}{4} \langle (O^1)^2 \rangle + \frac{1}{4} \langle (O^2)^2 \rangle + \frac{1}{2} \langle O^1 O^2 \rangle - (1 - 2\epsilon)^{2N} \text{Tr}(\bar{\Lambda}_Z(O)\rho^2)^2 \\ &= \frac{1}{2} \text{Tr}(\bar{\Lambda}_Z(O^2)\rho) + \frac{1}{2} \text{Tr}(\bar{\Lambda}_Z(O)\rho)^2 - (1 - 2\epsilon)^{2N} \text{Tr}(\bar{\Lambda}_Z(O)\rho^2)^2, \end{aligned} \quad (49)$$

where in the last step of Eq. (49) we have used the following fact to evaluate $\langle O^1 O^2 \rangle = \text{Tr}(\Lambda_{\text{tot}}^\dagger(O^1 O^2)\rho_{\text{in}})$

$$\begin{aligned} \Lambda_{\text{tot}}^\dagger(O^1 O^2) &= \circ_{k=1}^N (\mathcal{U}_k^\dagger \circ \tilde{\Lambda}_k^\dagger)(O^1 O^2) \\ &= |0\rangle\langle 0| \otimes \bar{\Lambda}_Z(O)^{\otimes 2} + |1\rangle\langle 1| \otimes \left(\prod_{k=1}^N \text{SWAP}_k \right) \bar{\Lambda}_Z(O)^{\otimes 2} \left(\prod_{k=1}^N \text{SWAP}_k \right) \\ &= |0\rangle\langle 0| \otimes \bar{\Lambda}_Z(O)^{\otimes 2} + |1\rangle\langle 1| \otimes \bar{\Lambda}_Z(O)^{\otimes 2} \\ &= I \otimes \bar{\Lambda}_Z(O)^{\otimes 2}, \end{aligned} \quad (50)$$

and hence $\text{Tr}(\Lambda_{\text{tot}}^\dagger(O^1 O^2)\rho_{\text{in}}) = \text{Tr}(|+\rangle\langle +|) \text{Tr}(\bar{\Lambda}_Z(O)\rho)^2 = \text{Tr}(\bar{\Lambda}_Z(O)\rho)^2$. Given these calculations we get that the variance of the estimator for dephasing errors in the virtual distillation circuit is

$$\begin{aligned} \text{Var}(\text{estim.}) &= \frac{1}{R(1 - 2\epsilon)^{2N}} \left(\frac{\text{Tr}(\bar{\Lambda}_Z(O^2)\rho)}{2 \text{Tr}(\rho^2)^2} + \frac{\text{Tr}(\bar{\Lambda}_Z(O)\rho)^2}{2 \text{Tr}(\rho^2)^2} + \frac{\text{Tr}(\bar{\Lambda}_Z(O)\rho^2)^2}{\text{Tr}(\rho^2)^4} \right. \\ &\quad \left. - 2 \frac{\text{Tr}(\bar{\Lambda}_Z(O)\rho^2)}{\text{Tr}(\rho^2)^3} \text{Tr}(\bar{\Lambda}_Z(O)\rho) \right). \end{aligned} \quad (51)$$

Comparing this equation with the expression for the noiseless variance of the estimator given by Eq. (16), we see that Eq. (51) is scaled by a factor $1/(1 - 2\epsilon)^{2N}$ and that O has been replaced by $\bar{\Lambda}_Z(O)$.

C Coherent mismatch

In this section we investigate the coherent mismatch, also known as drift, of the dominant eigenvector that is caused by the noise in the QAOA circuit. Since virtual distillation relies on the assumption that the dominant eigenvector approximates the ideal state, it is of importance to quantify how good of an approximation this is. The coherent mismatch is defined as [\[26\]](#)

$$c := 1 - |\langle \psi_{\text{dom}} | \psi_{\text{ideal}} \rangle|^2, \quad (52)$$

where $|\psi_{\text{dom}}\rangle$ is the dominant eigenvector of a noisy quantum state ρ and $|\psi_{\text{ideal}}\rangle$ is with respect to the ideal (noiseless) state. In [FIG. 5a](#) we plot the coherent mismatch, averaged over all instances, between the ideal noiseless QAOA state, and the dominant eigenvector of $\rho_\Lambda(\alpha_{\text{opt}}, \beta_{\text{opt}})$, where the optimal angles are with respect to $C_{\text{mitigated}}^\Lambda(\alpha, \beta)$. From the figure it can be seen that the coherent mismatch is smaller for dephasing errors in the QAOA circuit compared to both depolarizing errors and amplitude damping. For the case of depolarizing errors this fact can be explained by converting an incoherent error to a coherent one. Consider a Z error that occurs before the $X(\beta)$ -gate. By commuting the error through the gate we get

$$X(\beta)Z = e^{-i\beta X} Z = \cos(\beta)Z - i \sin(\beta)XZ = ZX(-\beta), \quad (53)$$

where in the last equality we have used the anti-commutation relation between X and Z , $XZ = -ZX$. Hence, we see that a Z error introduces an incoherent error that corresponds to an extra π rotation of the $X(\beta)$ -gate.

For the $ZZ(\alpha)$ -gate a Z error acting on any of the two-qubits commute through the gate, and hence only the $X(\beta)$ -gate will be affected by dephasing errors which causes coherent mismatch. For the depolarizing channel, however, both X and Y errors will lead to a coherent error of the $ZZ(\alpha)$ -gate which will result in a greater coherent mismatch. Nevertheless, we see that the coherent mismatch is small for large error probabilities, less than 2% for depolarizing error.

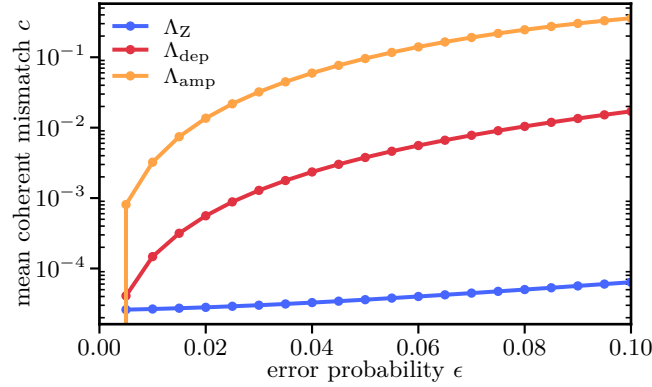


Figure 5: The coherent mismatch averaged over all instances. The blue line is the mean coherent mismatch with dephasing errors in the QAOA circuit and the red line is with depolarizing errors in the QAOA circuit.