Efficient Readout Error Mitigation Heuristic Using Singular Value Decomposition

Bo Yang[1] and Rudy Raymond[2]
[1] *Graduate School of Information Science and Technology, The University of Tokyo*
[2] *IBM Quantum, IBM Research Tokyo*
(Dated: July 13, 2021)

Nowadays, the near-term quantum devices are becoming available, and the results of quantum algorithms on such devices would be affected by many kinds of noise. One of the significant noise is the readout error, which can be mitigated by the classical post-processing. The basic process of quantum readout error mitigation (QREM) is to apply the inversion of calibration matrix describing the transition rates from the initial states to the measurement outcomes. However, the rigorous matrix inversion would take exponential time and memory to the size of qubit, while the qubit number of latest quantum devices are growing to over 50. In order to implement QREM for the results from the devices with tens or more qubits, we propose a polynomial time and memory heuristics to the number of qubits based on a bold assumption to the measurement outcomes. The idea of the proposed algorithm is to use singular value decomposition on the calibration matrix and efficiently approximate vectors in mitigation process. We also perform a numerical simulation on the modified Grover algorithm, which has wide application and the possibility to be run on near future quantum devices. Focusing on the property of the modified Grover algorithm, we checked the mitigated results by the proposed method shows almost the same performance to the rigorously mitigated one and outperformed the results by existing method.

I. INTRODUCTION

These years, many near-term quantum devices are open to public [1–7]. Such near-term devices are still too noisy and small to incorporate quantum error correction methods [8–11] for protecting quantum information. In order to obtain better results from the near-term quantum algorithms on the current and near future quantum devices, various methods for mitigating the errors have been proposed such as zero-noise extrapolation for mitigating whole errors [12–14], probabilistic error cancellation subject to a Markovian noise model [12, 15], dynamic decoupling for the decoherence errors [16, 17], and many new methods for various noises and models [18, 19].

One of the significant noise factors on current devices is the readout error occurring in the measurement process. The readout error can be characterized by a stochastic matrix called calibration matrix, whose elements represent the transition probability of the input states and the measurement outputs [20, 21]. The classical post-processing using the calibration matrix to reduce the effect of the readout errors from the probability vector of measurement outcome is called quantum readout error mitigation (QREM). The basic way of QREM is to apply the inverse of the calibration matrix to the noisy probability vector to recover the clean probability vector or the expectation values. Generally mitigating the mean values of observables seems to require less computational resources and many methods are proposed for different purposes [22, 23]. On the other hand, mitigating probability distribution itself would require more efforts. The rigorous inversion of calibration matrix to the probability vector would cost exponential time and memory to the number of measured qubits. Since the near-term devices with over 50 qubits has appeared and those with hundreds of qubits are expected to be realized in several years, existing QREM methods [5, 24, 25] are becoming not applicable to such large systems. However, while designing the scalable QREM methods for such large devices becomes a pressing issue, only few researches [26, 27] aim to tackle this as far as we know. Our goal is to propose a scalable QREM method for measurement results from large quantum systems. In addition, it is also required to output the physically proper probability vector which satisfy non-negativity and the sum of elements exactly to one. Since the direct matrix inversion would not meet these constrains, the open-source software package Qiskit [5] uses constrained least square optimization method, which would cost much longer time than the matrix inversion method.

Based on the above requirements, we propose a $O(ns^2)$ time and $O(s)$ memory heuristic algorithm to the number of qubit n and shot count s, introducing a bold assumption on probability vector. The assumption is that the zero valued elements in the noisy probability vector are ignorable. In other words, only non-negative elements are stored and used for the computation through the proposed method. This assumption is applicable when there are only few states that are expected to be measured. Besides, the proposed method works under the tensor product noise model, where the readout error on each qubit is assumed to be local. The proposed method consists of three steps. First, apply the inverse of calibration matrix to the focused elements under the assumption as mentioned above and output a roughly mitigated vector. Then, find the correction vector to be added to the roughly mitigated vector, making the sum of all elements in the vector become one. Finally, use the negative cancelling technique by Smolin, Gambetta, and Smith [28] to remove the negative values in the vector. The core of proposed method is to approximate the correction vector as a solution of constrained least square problem formulated by the singular value decomposition (SVD) of the calibration matrix. By adding further assumption on the readout error, this approximation process becomes much faster (see III 2). Compared to the method by Mooney et al. [26], the proposed method applies the inverse of calibration matrix correctly in the first step.

We also ran the numerical simulation of the proposed method on the quantum simulator provided by Qiskit [5]. The problem used here is the modified Grover algorithm by Uno et al. [29]. There are several motivations to focus on this algorithm. One is that the problem setting of this algorithm is likely to fit the assumption of our proposed QREM method in that most of the measured states are expected to be all-zero state: $|0\rangle$. Besides, the modified Grover algorithm is based on the maximum likelihood amplitude estimation (MLAE) algorithm [30] which avoid using phase estimation in the original amplitude estimation method [31]. Amplitude estimation has important applications in the field of finance and machine learning using quantum devices [32, 33]. By using less controlled gates and parallel running the shallower Grove iteration circuits, this type of amplitude estimation estimation algorithm is more expected to run on the near future devices on which phase estimation would not work. Regarding the modified Grover algorithm is more tolerant to depolarization errors, this numerical simulation suggests that this algorithm is also possible to be implemented under readout noises. Therefore this numerical simulation not only supports the scalability and the application possibility of proposed method but also implies the possibility of implementation of modified Grover algorithm on near future quantum devices.

II. PRELIMINARIES

A. Tensor Product Noise Model

Due to the leakage of the measurement pulse on qubit, measurement error may have correlation with the qubits around. However, in this work, we discuss the readout noise under the situation where the readout errors are either local or correlated among limited spatial extent. Under this model, the calibration matrix A with size $2^n \times 2^n$ to the number of qubit n can be seen as a tensor product of small fractions of calibration matrices for local areas. This can much reduce the number of measurements for preparing the calibration matrix and the space for storing the calibration matrix. On the current near-term devices of IBM Quantum Experience, this model is shown to be valid, representing the local error occurs on each qubit is dominant [26]. Hereinafter, we assume the measurement error is completely local, that is, $A := \bigotimes_{i=0}^{n} A_i$ where A_i is the 2×2 calibration matrix of qubit i.

B. Assumption

We also introduce an assumption to the whole QREM process, that the unmeasured elements in the noisy probability vector are ignorable. By this assumption, the maximum memory to store the probability vector becomes the order of finite shot count s. This assumption may hold true when only few states are ideally expected to be measured without errors because otherwise the true state to be mitigated would not appear in the noisy probability vector. Although this may be too bold, we will see in the numerical simulation that there does exist the case that the assumption works effectively. Besides, it is also possible to extend range of states to the elements with small Hamming distances to the measured states. Hereinafter, let the subscript $_S$ denote the limitation of the vector to the set of elements in the set S. (e.g. For the vector x, the limitation of its elements to the set S is denoted by x_S.)

C. Problem Setting

We are considering the following problem. We are given a probability vector $y \in \mathbb{R}^N$ as an n-qubit measurement outcome such that $y \geq 0$ and $1^T \cdot y = \sum_j y_j = 1$ where $N = 2^n$. We can also get access to the calibration matrix A_i for each qubit i. Then the task is to recover the *true* probability vector $\tilde{x} \in \mathbb{R}^N$ before the noisy channel of measurement. This problem can be formulated as optimization problem to find the closest vector satisfying the two constraints above. In this work, the least square method is adopted as existing researches and packages are widely using [5, 21]. Since we employ the assumption (II B), the problem formulation becomes

$$\underset{\tilde{x}_S}{\text{minimize}} \quad \|A\tilde{x}_S - y\|^2 \tag{1}$$

$$\text{subject to} \quad 1^T \cdot \tilde{x}_S = 1 \tag{2}$$

$$\tilde{x}_S \geq 0 \tag{3}$$

III. PROPOSED ALGORITHMS

To solve the defined optimization problem efficiently, we will take three steps as mentioned in the introduction part. Let the set of positions of non-zero elements in noisy probability vector y be S. First, apply the inverse of calibration matrix to the focused elements under the assumption as mentioned above and output a roughly mitigated vector x_S. Then, find the correction vector Δ_S which makes x_S satisfy $1^T \cdot \hat{x}_S = 1$ where $\hat{x}_S = x_S + \Delta_S$. Finally, cancel the negative values in the corrected vector $\hat{x}_S$.

1. Step 1: Matrix Inverse

Let $x \in \mathbb{R}^{2^n}$ be $x := A^{-1}y$. Since calibration data are given by n 2×2 matrices, each element of x can be computed in $O(ns)$ time respectively by exploiting the property of tensor product of matrices. The algorithm is described at 1. While computing all elements of x requires $O(sn2^n)$ time and $O(2^n)$ memory, computing only the elements in S, the computational cost would be $O(ns^2)$ time and $O(s)$ memory. Let $x_S \in \mathbb{R}^{|S|}$ denote the restricted vector of x into the elements of S. Computing x_S is the rate-determining step throughout whole processes in the proposed method.

2. Step 2: Making All Element Sum to One

Next is to find a correction vector Δ_S that makes the element sum of the vector to one. To compute Δ_S, we first consider the case when the assumption II B is not applied. Let $\hat{x}$ be $\hat{x} = x + \Delta$. Then Δ is approximated based on the following least square problem.

$$\underset{\Delta}{\text{minimize}} \quad \|A\hat{x} - y\|^2 = \|A\Delta\|^2 \tag{4}$$

$$\text{subject to} \quad 1^T \cdot \hat{x} = 1 \tag{5}$$

The idea is to perform the singular value decomposition (SVD) of A and convert the optimization problem above to the form which is analytically solvable. Let the SVD of A be $A = U\Sigma V^T = \sum_{i=0}^{N-1} \sigma_i u_i v_i^T$ and represent Δ as $\Delta = \Delta_{j_0} v_{j_0} + \Delta_{j_1} v_{j_1} + \cdots \Delta_{j_{k-1}} v_{j_{k-1}}$ using k right singular vectors $\{v_i\}$ of A. Then the problem (4) becomes

$$\begin{aligned} &\min_{\Delta \in \mathbb{R}^N} \sum_{i=j_0}^{j_{k-1}} \sigma_i^2 \Delta_i^2 \\ &\text{subject to} \sum_{i=j_0}^{j_{k-1}} \left(1^T \cdot v_i\right) \Delta_i = 1 - 1^T \cdot x \ . \end{aligned} \tag{6}$$

This constrained least square problem can be rigorously solved by Lagrange multiplier. Each coefficient of $\Delta = \Delta_{j_0} v_{j_0} + \Delta_{j_1} v_{j_1} + \cdots \Delta_{j_{k-1}} v_{j_{k-1}}$ can be computed as

$$\Delta_i = \frac{1 - 1^T \cdot x}{\sum_{l=j_0}^{j_{k-1}} \frac{\left(1^T \cdot v_l\right)^2}{\sigma_l^2}} \cdot \frac{1^T \cdot v_i}{\sigma_i^2}. \tag{7}$$

Since the calibration matrix A is the tensor product of small matrix A_i for each qubit i, the values $\sigma_i, 1^T \cdot v_i$ can be computed in $O(n)$ time using the property of $(U_1\Sigma_1 V_1^T) \otimes (U_2 \Sigma_2 V_2^T) = (U_1 \otimes U_2)(\Sigma_1 \otimes \Sigma_2)(V_1^T \otimes V_2^T)$ for the SVD of two matrices $A_1 = U_1 \Sigma_1 V_1^T$ and $A_2 = U_2 \Sigma_2 V_2^T$. Applying the assumption II B, the correction vector is approximated as Δ_S. The time complexity to compute the coefficients Δ_i is $O(nsk)$ with arbitrary parameter k.

Furthermore, when the readout error probability $p(1|0)$ and $p(0|1)$ are almost equal, the coefficients of Δ can be more efficiently approximated. The idea is that, when $p(1|0) \simeq p(0|1)$, the calibration matrix of each qubit A_i is close to symmetric matrix, which can be eigen decomposed by Hadamard matrices. This idea can be stated as the following lemma. Finally the coefficient Δ_i under this additional assumption can be computed as

$$\Delta_i' = \frac{1 - 1^T \cdot x}{\frac{(1^T \cdot v_0)^2}{\sigma_0^2}} \cdot \frac{1^T \cdot v_i}{\sigma_i^2} \ . \tag{8}$$

In addition, (8) implies $\Delta_0' \gg \Delta_i'$ for $i = 1, 2, \cdots$. We used $\Delta_S' = \Delta_0' v_{0S}$ for the correction vector in the numerical simulation.

3. Step 3: Negative Cancelling

Finally, we are going to delete the negative values in $\hat{x}$. In this process, we can use the algorithm by Smolin, Gambetta, and Smith [28] which assumes the input vector satisfy the condition (2) (hereinafter, this is called "SGS algorithm"). The procedure of SGS algorithm is described at the algorithm 2. Through SGS algorithm, the finally mitigated probability vector $\tilde{x}$ (= sgs_algorithm($\hat{x}$)) is computed in $O(N \log N)$ time where N is the number of elements in $\hat{x}$. Due to the assumption II B, $N = |S| < s$ and the computational time is $O(s \log s)$.

IV. NUMERICAL SIMULATIONS

A. Experiment Settings

We conduct a Monte Carlo integration by modified Grover algorithm [29] to study the performance of proposed method. This also aimed to check the existence of important applications of the proposed method. The method by Mooney et al. [26] is also provided and tested to compare the results with the proposed method.

First, we roughly review the process of modified Grover algorithm. The whole procedure to estimate the amplitude follows the MLAE algorithm [30], running shallower quantum circuits with different iterations of the Grover operator. The modified Grover algorithm only changes the Grover operator from MLAE algorithm. The operator is represented as $Q = U_0 A^\dagger U_f A$, where U_0 and U_f are the reflection operators defined as

$$\begin{aligned} U_0 &= -\mathbf{I}_{n+1} + 2|0\rangle_{n+1} \langle 0|_{n+1} \, , \\ U_f &= -\mathbf{I}_{n+1} + 2\mathbf{I}_n \otimes |0\rangle\langle 0| . \end{aligned} \tag{9}$$

The initial state $|0\rangle_{n+1}$ after m iterations of operator Q becomes

$$Q^m |0\rangle_{n+1} = \cos(2m\theta) |0\rangle_{n+1} + \sin(2m\theta) |\phi\rangle_{n+1} \tag{10}$$

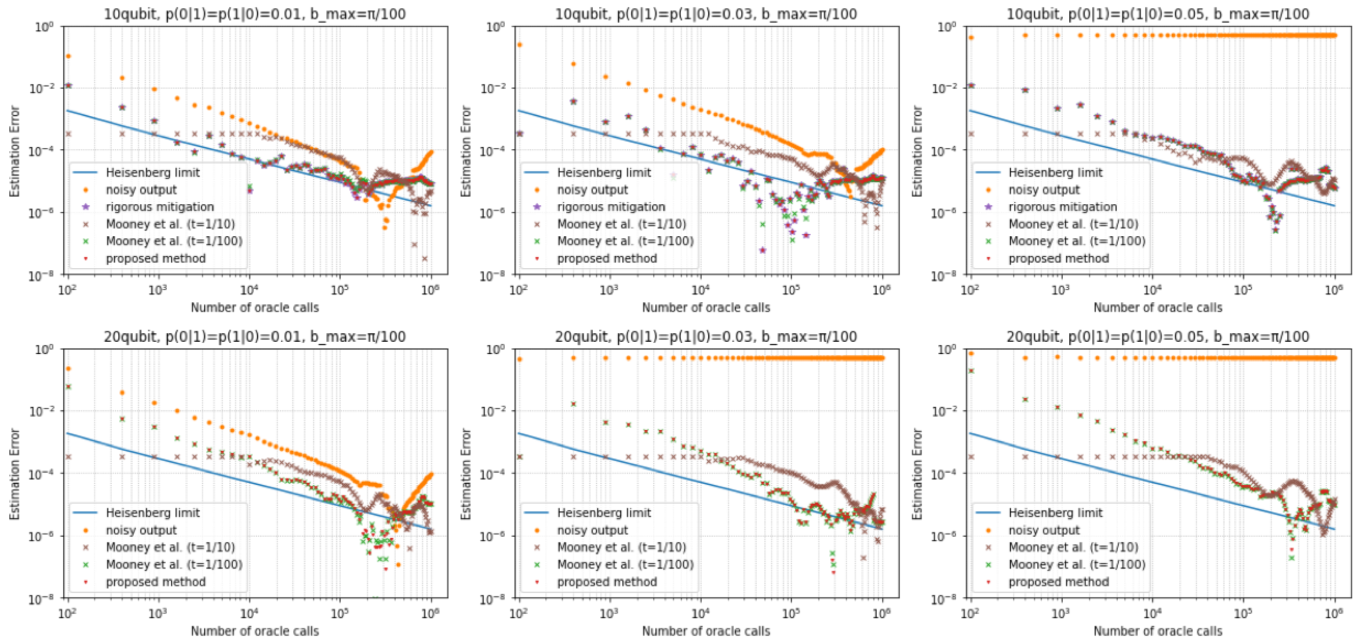

FIG. 1. The estimation error of Monte Carlo integration on 10-qubit and 20-qubit system. The blue lines are the lower bound of estimation error in terms of the Fisher information. The orange plots are the results of noisy circuit simulation with readout error. The purple label "rigorous mitigation" means the mitigated results by applying rigorous matrix inverse. The brown label "Mooney et al. (t=1/10)" represents the results by the QREM method of Mooney et al. [26] with its threshold set by $t = 1/10$. The green label "Mooney et al. (t=1/100)" represents the results by the QREM method of Mooney et al. [26] with its threshold set by $t = 1/100$. The red label "proposed method" describes the estimation errors by applying the proposed QREM method.

where $|\phi\rangle_{n+1} \neq |0\rangle_{n+1}$. It is enough to know the probability of getting state $|0\rangle_{n+1}$ in 10, to estimate the value θ. According to [30], the estimation errors would follow the Heisenberg limit that the error decreases in the speed of $O(1/m)$ for m rounds of Grover iterations. The performance of the modified Grover algorithm can be checked by seeing how well the estimation errors follows the Heisenberg limit.

We applied this modified Grover algorithm to Monte Carlo integration, following the procedures in [30]. Using the notations in [30], the goal of the Monte Carlo integration is to compute the following value

$$\begin{aligned} I &= \frac{1}{b_{\max}} \int_0^{b_{\max}} \sin(x)^2 dx \\ &= \frac{1}{b_{max}} \left(\frac{b_{max}}{2} - \frac{1}{4} \sin(2b_{max}) \right) \end{aligned} \tag{11}$$

where b_{max} is an arbitrary constant. This value can be estimated via the modified Grover operations above and we numerically checked the gap of estimation error to the Heisenberg limit.

The numerical simulation was performed with 10-qubit and 20-qubit search space respectively on the Qiskit simulator [5]. Here we run the modified Grover operators for $m = 0, 1, 2, \cdots, 99$ times with $b_{max} = \pi/100$. Since our algorithm scales quadratic to the number of shots, the number of shots is set to $N_{shot} = 100$. Besides, we tested different readout error rates of $p(0|1) = p(1|0) = 0.01, 0.03, 0.05$. These parameters used in the numerical simulation are listed in Table I.

TABLE I. List of parameters used in the numerical simulation.

number of qubits	n	$\{10, 20\}$
number of shots	N_{shot}	100
number of Grover iteration	m	$\{0, 1, 2, \cdots, 99\}$
target values	$I = \cos^2\theta$	$b_{max} = \pi/100$
readout noise	$p(0|1) = p(1|0)$	$\{0.01, 0.03, 0.05\}$

B. Results

The results of the numerical simulation are shown in Fig. 1. For the 10-qubit system, we can see the estimation errors of mitigated plots are more closed to the Heisenberg limit than the plots containing readout error for all readout error rates. In the case of $p(0|1) = p(1|0) = 0.05$, the readout error might become so large that the plot with readout errors (the orange dots) is not likely to follow the Heisenberg limit.

The purple, green, and red plots are almost at the same position on each figure. Since the the purple plot means the standard method through the rigorous inverse calibration matrix, our algorithm can be seen as good approximation to such standard way.

For the 20-qubit system, we can also see the estimation errors of mitigated plots are well following the Heisenberg limit, while the noisy plots colored by orange are not following it at the readout error rate 0.03 and 0.05. This means one need readout error mitigation to obtain the advantageous results from modified Grover algorithm. In the size of 20-qubit, the conventional exponential time methods would not finish in practical time. On the other hand, our algorithm still outputs the plots which follows the Heisenberg limits. Since the results by conventional method and the results of our methods are very closed for 10-qubit system, we guess the output by conventional methods for 20-qubit system would also closed to the results by our algorithms, meaning the potential advantage of our algorithm.

V. CONCLUSION

Our algorithm mitigates the readout error with $O(ns^2)$ time and $O(s)$ memory with n qubits and s shots through the post-processing on classical computers. Fixing the number of shots, our algorithm scales linearly to the number of measured qubits, which provides a scalable readout error mitigation tool for the current and near future quantum devices with larger qubits. Besides, every step in our algorithm can be implemented in parallel to the vector elements. Therefore, our algorithm has the potential to get more accelerated with the help of GPU.

The numerical simulation on the modified Grover algorithm shows our algorithm has good performance even on the system with larger qubits towards the mitigation of modified Grover algorithm with readout errors. The the estimation error of mitigated results on 20-qubit system still follows the Heisenberg limit, which might support the advantage of our algorithm when the size of qubit becomes large.

Some points in our current algorithm are to be further improved and those are still work in progress. First, the precision of our current algorithm is not theoretically bounded due to the assumption we set in our algorithm that the labels appearing in the mitigated vector always appear in the measured probability vector. The remaining task includes the analysis of the closeness between the mitigated probability distribution and the ideal probability distribution. In addition, this assumption also restricts the application range of our algorithm to the measurement result with only few labels in expected outcomes because there might be much more source labels than those are measured if the number of shots are much smaller than the number of possible state labels. To better utilize our proposed algorithm, one can find more practical cases where this assumption is valid to be applied. Furthermore, one could also consider other assumptions such as the situation where the measured state string contains at most one or two errors.

Applying our algorithm to a different noise model is also a possible future direction. The tensor product noise model can only capture the local measurement error although local error is dominant in the current near-term quantum devices [26]. Instead, the correlated Markovian noise model may include cross-talk errors. Using the variation of this noise model called continuous time Markov processes (CTMP), Bravyi et al. [22] have recently developed more efficient readout error mitigation protocol to directly compute expectation values form measured probability vectors. Since our algorithm aims to correct the probability vector itself, one might incorporate the method in [22] to develop our algorithm to the CTMP noise model.

[1] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, A. Kandala, G. A. Keefe, K. Krsulich, W. Landers, E. P. Lewandowski, D. T. McClure, G. Nannicini, A. Narasgond, H. M. Nayfeh, E. Pritchett, M. B. Rothwell, S. Srinivasan, N. Sundaresan, C. Wang, K. X. Wei, C. J. Wood, J.-B. Yau, E. J. Zhang, O. E. Dial, J. M. Chow, and J. M. Gambetta, Demonstration of quantum volume 64 on a superconducting quantum computing system (2020), arXiv:2008.08571 [quant-ph].

[2] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. S. Allman, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, and B. Neyenhuis, Demonstration of the qccd trapped-ion quantum computer architecture (2020), arXiv:2003.01293 [quant-ph].

[3] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, D. Eppens, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, L. B. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. N. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, M. Leib, E. Lucero, O. Martin, J. M. Martinis, J. R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, F. Neukart, H. Neven, M. Y. Niu, T. E. O'Brien, B. O'Gorman, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, A. Skolik, V. Smelyanskiy, D. Strain, M. Streif, K. J. Sung, M. Szalay, A. Vainsencher, T. White, Z. J. Yao, P. Yeh, A. Zalcman, and L. Zhou, Quantum approximate optimization of non-planar graph problems on a planar superconducting processor (2020), arXiv:2004.04197 [quant-ph].

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, and

et al., Science **369**, 1084–1089 (2020).

[5] G. Aleksandrowicz, T. Alexander, P. Barkoutsos, L. Bello, Y. Ben-Haim, D. Bucher, F. J. Cabrera-Hernández, J. Carballo-Franquis, A. Chen, C.-F. Chen, J. M. Chow, A. D. Córcoles-Gonzales, A. J. Cross, A. Cross, J. Cruz-Benito, C. Culver, S. D. L. P. González, E. D. L. Torre, D. Ding, E. Dumitrescu, I. Duran, P. Eendebak, M. Everitt, I. F. Sertage, A. Frisch, A. Fuhrer, J. Gambetta, B. G. Gago, J. Gomez-Mosquera, D. Greenberg, I. Hamamura, V. Havlicek, J. Hellmers, Łukasz Herok, H. Horii, S. Hu, T. Imamichi, T. Itoko, A. Javadi-Abhari, N. Kanazawa, A. Karazeev, K. Krsulich, P. Liu, Y. Luh, Y. Maeng, M. Marques, F. J. Martín-Fernández, D. T. McClure, D. McKay, S. Meesala, A. Mezzacapo, N. Moll, D. M. Rodríguez, G. Nannicini, P. Nation, P. Ollitrault, L. J. O'Riordan, H. Paik, J. Pérez, A. Phan, M. Pistoia, V. Prutyanov, M. Reuter, J. Rice, A. R. Davila, R. H. P. Rudy, M. Ryu, N. Sathaye, C. Schnabel, E. Schoute, K. Setia, Y. Shi, A. Silva, Y. Siraichi, S. Sivarajah, J. A. Smolin, M. Soeken, H. Takahashi, I. Tavernelli, C. Taylor, P. Taylour, K. Trabing, M. Treinish, W. Turner, D. Vogt-Lee, C. Vuillot, J. A. Wildstrom, J. Wilson, E. Winston, C. Wood, S. Wood, S. Wörner, I. Y. Akhalwaya, and C. Zoufal, Qiskit: An Open-source Framework for Quantum Computing (2019).

[6] P. J. Karalekas, N. A. Tezak, E. C. Peterson, C. A. Ryan, M. P. da Silva, and R. S. Smith, Quantum Science and Technology **5**, 024003 (2020).

[7] 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, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Nature **574**, 505 (2019).

[8] P. W. Shor, Phys. Rev. A **52**, R2493 (1995).

[9] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).

[10] A. Kitaev, Annals of Physics **303**, 2–30 (2003).

[11] C. Chamberland, G. Zhu, T. J. Yoder, J. B. Hertzberg, and A. W. Cross, Physical Review X **10**, 10.1103/physrevx.10.011022 (2020).

[12] K. Temme, S. Bravyi, and J. M. Gambetta, Physical Review Letters **119**, 10.1103/physrevlett.119.180509 (2017).

[13] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Nature **567**, 491 (2019).

[14] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, 2020 IEEE International Conference on Quantum Computing and Engineering (QCE) 10.1109/qce49297.2020.00045 (2020).

[15] R. Takagi, Optimal resource cost for error mitigation (2020), arXiv:2006.12509 [quant-ph].

[16] K. X. Wei, I. Lauer, S. Srinivasan, N. Sundaresan, D. T. McClure, D. Toyli, D. C. McKay, J. M. Gambetta, and S. Sheldon, Phys. Rev. A **101**, 032343 (2020).

[17] B. Pokharel, N. Anand, B. Fortman, and D. A. Lidar, Physical Review Letters **121**, 10.1103/physrevlett.121.220502 (2018).

[18] J. Sun, X. Yuan, T. Tsunoda, V. Vedral, S. C. Benjamin, and S. Endo, Physical Review Applied **15**, 10.1103/physrevapplied.15.034026 (2021).

[19] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, Nature Communications **11**, 10.1038/s41467-020-14341-w (2020).

[20] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, Nature Physics **5**, 27 (2009).

[21] F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, Quantum **4**, 257 (2020).

[22] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, Phys. Rev. A **103**, 042605 (2021).

[23] G. S. Barron and C. J. Wood, Measurement error mitigation for variational quantum algorithms (2020), arXiv:2010.08520 [quant-ph].

[24] B. Nachman, M. Urbanek, W. A. de Jong, and C. W. Bauer, npj Quantum Information **6**, 10.1038/s41534-020-00309-7 (2020).

[25] R. Hicks, C. W. Bauer, and B. Nachman, Phys. Rev. A **103**, 022407 (2021).

[26] G. J. Mooney, G. A. L. White, C. D. Hill, and L. C. L. Hollenberg, Generation and verification of 27-qubit Greenberger-Horne-Zeilinger states in a superconducting quantum computer (2021), arXiv:2101.08946 [quant-ph].

[27] E. Peters, A. C. Y. Li, and G. N. Perdue, Perturbative readout error mitigation for near term quantum computers (2021), arXiv:2105.08161 [quant-ph].

[28] J. A. Smolin, J. M. Gambetta, and G. Smith, Physical Review Letters **108**, 10.1103/physrevlett.108.070502 (2012).

[29] S. Uno, Y. Suzuki, K. Hisanaga, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto, Modified grover operator for amplitude estimation (2020), arXiv:2010.11656 [quant-ph].

[30] Y. Suzuki, S. Uno, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto, Quantum Information Processing **19**, 10.1007/s11128-019-2565-2 (2020).

[31] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, Quantum Computation and Information , 53–74 (2002).

[32] T. Giurgica-Tiron, I. Kerenidis, F. Labib, A. Prakash, and W. Zeng, Low depth algorithms for quantum amplitude estimation (2020), arXiv:2012.03348 [quant-ph].

[33] A. Bouland, W. van Dam, H. Joorati, I. Kerenidis, and A. Prakash, Prospects and challenges of quantum finance (2020), arXiv:2011.06492 [q-fin.CP].

Appendix A: Computing the Matrix Inversion Element-wise

Algorithm 1 Mitigate One State (mitigate_one_state)

Require: target state label t, probability vector y
Ensure: mitigated count c
  $c \leftarrow 0$
  for source state label s in y **do**
    product $p \leftarrow 1$
    for calibration matrix A^k in $\{A^0, \cdots, A^{n-1}\}$ **do**
      $i \leftarrow$ the value of k-th digit of t (0 or 1)
      $j \leftarrow$ the value of k-th digit of s (0 or 1)
      $p \leftarrow p \cdot A^k_{i,j}$
    end for
    $c \leftarrow c + p \cdot y[s]$
  end for
  return c

Appendix B: The Negative Cancelling algorithm by Smolin, Gambetta, and Smith

---

Algorithm 2 Negativity Cancellation by Smolin, Gambetta, and Smith [28] (sgs_algorithm)

Require: vector $\hat{\mathbf{x}}$ (satisfying $\mathbf{1}^T\hat{\mathbf{x}} = 1$)
Ensure: mitigated probability vector $\tilde{\mathbf{x}}$
  $queue \leftarrow$ make a priority queue
  accumulator of positive values $p \leftarrow 0$
  for state label s in $\mathbf{y}$ **do**
    if $\hat{\mathbf{x}}[s] > 0$ **then**
      $queue.push(\hat{\mathbf{x}}[s])$
      $p \leftarrow p + \hat{\mathbf{x}}[s]$
    end if
  end for
  accumulator of negative values $neg \leftarrow 1 - p$
  while $queue$ is not empty **do**
    if $(queue.top() + neg)/queue.size() < 0$ **then**
      $neg \leftarrow queue.pop()$
    else
      break
    end if
  end while
  mitigated counts $\tilde{\mathbf{x}} \leftarrow$ empty dictionary
  division value of negative accumulator $d \leftarrow queue.size()$
  while $queue$ is not empty **do**
    $\tilde{\mathbf{x}}[s] \leftarrow \hat{\mathbf{x}}[s] + neg/d$
  end while
  return $\tilde{\mathbf{x}}$
