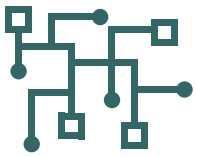


ISE

Industrial and
Systems Engineering

# A Bayesian Approach for Characterizing and Mitigating Gate and Measurement Errors


MUQING ZHENG[1], ANG LI[2], TAMÁS TERLAKY[1], AND XIU YANG[1]

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

[2]Advanced Computing, Mathematics and Data Division, Pacific Northwest National Laboratory, Richland, WA 99354



ISE Technical Report 20T-028


LEHIGH
UNIVERSITY

# A Bayesian Approach for Characterizing and Mitigating Gate and Measurement Errors

Muqing Zheng[1], Ang Li[2], Tamás Terlaky[1], and Xiu Yang[1]

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015

[2]Advanced Computing, Mathematics and Data Division, Pacific Northwest National Laboratory, Richland, WA 99354



**Abstract**

Various noise models have been developed in quantum computing study to describe the propagation and effect of the noise from imperfect implementation of hardware. In these models, critical parameters, e.g., error rate of a gate, are typically modeled as constants. Instead, we model such parameters as random variables, and apply a new Bayesian inference algorithm to classical gate and measurement error models to identify the distribution of these parameters. By charactering the device errors in this way, we futher improve error filters accordingly. Experiments conducted on IBM's quantum computing devices suggest that our approach provides better error-mitigation performance than existing error-mitigation techniques, in which error rates are estimated as deterministic values. Our approach also outperforms the standard Bayesian inference method in such experiments.

## 1 Introduction

While quantum computing (QC) displays an exciting potential on reducing the time complexity of various problems, the noise from environment and hardware may still undermine the advantages of QC algorithms [17]. In the noisy intermediate-scale quantum (NISQ) era [17], error mitigation techniques have been attracted much attention, e.g., [9, 19, 7, 11, 10, 14, 4], as their cost is much lower than the quantum error correction codes [13, 3, 12, 20] in terms of circuit depth and number of qubits. One of the directions in the error mitigation study is to correct the measurement errors. These errors are usually modeled by applying a transition matrix on a probability vector, as such to depict the influence of the noise on the output of QC algorithms. More septically, the probability vector represents the output without measurement noise, the transition matrix describes how the noise affects this output, and the resulting vector consists of the probabilities of observing all possible states in practice. Here, the transition matrix can be constructed from conditional probabilities if only classical errors are considered, or from results of tomography if non-classical errors are also significant [6, 10, 14, 4]. On the other hand, the study in [18] shows the possibility to simulate bit-flip gate error in some quantum circuit in a classical manner.

The goal of error mitigation from algorithmic perspective is to recover the noise-free information using data from repeated experiments, which is usually achieved via statistical methods. In the existing error models, the parameters, e.g., error rate of measurement or gates, are usually considered as deterministic values. instead of, some random variables according to some unknown distributions. We did not see the related work for the latter case. By considering error mitigation as a stochastic inverse problem, we adopt a new Bayesian algorithm from [5] to construct the distributions of model parameters and use corresponding backward error models to filter out gate errors and measurement errors from the measurement outcomes of quantum circuits.

The paper is organized as following. In Section 2, we provide the measurement error model based on independent classical measurement error and expand the gate error model in [18] to multiple-error scenario. In Section 3, we introduce the use of Bayesian algorithm in [5] to infer the distributions of parameters of measurement error and gate error models. Then, we demonstrate the creation of our error

filter on IBM's quantum device `ibmqx2` (Yorktown) and apply our filter together with other existing error mitigation methods on measurement outcome from state tomography, an example of Grover's search [1], Quantum Approximate Optimization Algorithm (QAOA) [1], and a 200-NOT-gate circuit in Section 4.

Throughout this paper, we assume no state-preparation error.

# 2 Error Models

## 2.1 Notations

The goals of error models include estimating the influence of bit-flip gate error and measurement error in the outputs of a quantum circuit without accessing any quantum device and recovering the error-free output In this study, we assume that each gate has independent depolarizing error and each measurement operator has independent bit-flip error. We use three parameters to characterize the probabilities of the occurrence of di□erent types of errors as follows:

1. $\epsilon_g$ = the chance of having depolarizing error in a gate, where $\epsilon_g$ $(0, 1)$;
2. $\epsilon_{m0}$ = the chance of having measurement error when measure $|0\rangle$, where $\epsilon_{m0}$ $(0, 1)$;
3. $\epsilon_{m1}$ = the chance of having measurement error when measure $|1\rangle$, where $\epsilon_{m1}$ $(0, 1)$.

In 1-qubit case, the occurrence of depolarizing error is equivalent to the scenario that the qubit has $1 - \frac{3}{4}\epsilon_g$ chance to be error-free and $\frac{1}{4}\epsilon_g$ chance to have only bit-flip or phase-flip error, or the combination of both [15, p. 379]. We use Pauli matrices $X$, $Z$ and $Y$ to characterize the bit-flip error, the phase-flip error, and the combination of both bit-flip and phase-flip error, respectively. The depolarizing error $\epsilon_g$ build a connection between the probability of the occurrence of a general gate error and that of a bit-flip error. The approach of constructing the gate error model using $\epsilon_g$ in Section 2.3 is also applicable to other types of error models if such connection exists.

## 2.2 Measurement Error

### 2.2.1 Independent Measurement Errors

We assume that measurement errors are independent across qubits. As is demonstrated in [6], classical measurement error is applicable in the device we conduct experiments on, i.e., `ibmqx2`. We build measurement error model using conditional probabilities. Consider a single qubit with state $\alpha|0\rangle + \beta|1\rangle$, we have

$$\begin{aligned}\Pr(\text{Measure 0 w/ noise}) &= |\alpha|^2 \cdot (1 - \epsilon_{m0}) + |\beta|^2 \cdot \epsilon_{m1}\\ \Pr(\text{Measure 1 w/ noise}) &= |\alpha|^2 \cdot \epsilon_{m0} + |\beta|^2 \cdot (1 - \epsilon_{m1}),\end{aligned}$$

which is equivalent to

$$\begin{bmatrix} 1-\epsilon_{m0} & \epsilon_{m1} \\ \epsilon_{m0} & 1-\epsilon_{m1} \end{bmatrix} \begin{pmatrix} \Pr(\text{Measure 0 w/o noise}) \\ \Pr(\text{Measure 1 w/o noise}) \end{pmatrix} = \begin{pmatrix} \Pr(\text{Measure 0 w/ noise}) \\ \Pr(\text{Measure 1 w/ noise}) \end{pmatrix}, \tag{1}$$

where "w/" stands for with and "w/o" stands for without. Denoting $\epsilon_{m0}$ and $\epsilon_{m1}$ for qubit $i$ as $\epsilon_{m0,i}$ and $\epsilon_{m1,i}$, respectively, we can extend the matrix form in Eq. (1) to $n$-qubit cases as

$$Ar = \tilde{r}, \tag{2}$$

where

$$A := \bigotimes_{i=1}^{n} \begin{bmatrix} 1-\epsilon_{m0,i} & \epsilon_{m1,i} \\ \epsilon_{m0,i} & 1-\epsilon_{m1,i} \end{bmatrix}, \tag{3}$$

$$r := \begin{pmatrix} \Pr(\text{Measure 0...00 w/o noise}) \\ \Pr(\text{Measure 0...01 w/o noise}) \\ \vdots \\ \Pr(\text{Measure 1...11 w/o noise}) \end{pmatrix}, \tag{4}$$

$$\tilde{r} := \begin{pmatrix} \Pr(\text{Measure 0...00 w/ noise}) \\ \Pr(\text{Measure 0...01 w/ noise}) \\ \vdots \\ \Pr(\text{Measure 1...11 w/ noise}) \end{pmatrix}, \tag{5}$$

$A_{ij}$ [0,1], $r_i$ [0,1], $\tilde{r}_i$ [0,1]. We aim to identify $r$, but, in practice, we only have $\tilde{r}$ which is the probability vector characterizing the observed results from repeated experiments. Note that $A$ is a nonnegative left stochastic matrix (i.e., each column sums to 1), so if $r \geq 0$ and its entries sums to 1, $\tilde{r} \geq 0$ and its entries also sums to 1.

### 2.2.2 Measurement Error Filter

If $\epsilon_{m0,i}$ and $\epsilon_{m1,i}$ for all $i = 1, ..., n$ are known, the most straightforward denoising method derived from (2) is $r := A^{-1}\tilde{r}$. In practice, we usually can assume $\epsilon_{m0,i} < 0.5, \epsilon_{m1,i} < 0.5$ for all $i = 1, ..., 2^n$, so each individual 2-by-2 matrix has non-zero determinant. Thus $A$ has non-zero determinant and $A^{-1}$ exists. However, it is not guaranteed that $r \geq 0$. Alternative, one can compute

$$r := \underset{\sum_{i=1}^{2^n} r_i = 1,\ i \ 1,...,2^n\ r_i \geq 0}{\arg\min} \quad A^{-1}\tilde{r} - r \ _2, \tag{6}$$

and use $r$ to approximate $r$.

## 2.3 Bit-flip Error

We only consider the situation when phase-flip error does not a□ect a QC algorithm's outcomes. In this case, $Y$ error is equivalent to $X$ error. Together with original bit-flip error, for a 1-qubit gate with depolarizing error rate $\epsilon_g$, the probability of the occurrence of a bit-flip error is $\frac{1}{2}\epsilon_g$ [15]. Moreover, we will extend the error modeling approach to an $n-$qubit gate case, and we will demonstrate that $\frac{1}{2}\epsilon_g$ is still our major focus as in the 1-qubit scenario. The factor $\frac{1}{2}$ was introduced in the original literature [15], and it will not a□ect the validity of our conclusion.

Figure 1 illustrates a typical circuit we to study a single bit-flip error model. Here, $U$ is a noisy-free unitary gate and $X_{\epsilon_g/2}$ represents a bit-flip error with probability $\frac{1}{2}\epsilon_g$.

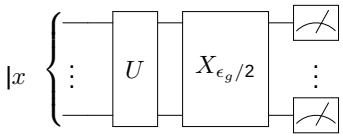



Figure 1: Single bit-flip error model

### 2.3.1 Single Bit-flip Error

We adopt the error model in [18] to describe the single bit-flip error. Of note, [18] does not provide a proof on the validity of this model, we complete the proof and extend it to multiple-error case in Section 2.3.2. Using this method, we first consider the case when there is only one gate and one bit-flip

error (on all qubits) in the circuit, as shown in Figure 1. Let $p : \{0,1\}^n \to [0,1]$, where $n$ is the number of qubits, be the Boolean function that represents the theoretical probability distribution of measurement outcomes and $x \in \{0,1\}^n$ denote the basis used in a QC algorithm. We perform Fourier expansion of this Boolean function as

$$p(x) = \sum_{s \in \{0,1\}^n} \hat{p}(s)(-1)^{s.x}, \tag{7}$$

where $\hat{p}(s)$ is Fourier coefficient of $p$ and $s.x = \sum_{i=1}^{n} s_i \cdot x_i$ [16, p.22]. These Fourier coefficients can be computed as

$$\hat{p}(s) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} p(x)(-1)^{s.x}.$$

Let $y$ be the erroneous version of $x$ induced by the bit-flip error, i.e.,

$$y_i = \begin{cases} x_i & \text{with probability } 1 - \frac{\epsilon_g}{2} \\ \neg x_i & \text{with probability } \frac{\epsilon_g}{2} \end{cases} \text{ for } i = 1, ..., n.$$

We use $\tilde{p} : \{0,1\}^n \to [0,1]$ to denote the distribution function of measurement outcomes under the noise model. Then (7) implies

$$\tilde{p}(x) = E_y[p(y)] = \sum_{s \in \{0,1\}^n} \hat{p}(s) E_y[(-1)^{s.x}].$$

Let us now only look at $E_y[(-1)^{s.x}]$. It is not hard to see

$$\begin{aligned} E_y[(-1)^{s.x}] &= E_y\left[\prod_{i=1}^{n} (-1)^{s_i \cdot y_i}\right] \\ &= \prod_{i=1}^{n} E_y\left[(-1)^{s_i \cdot y_i}\right] \\ &= \prod_{i=1}^{n} \left[\left(1 - \frac{\epsilon_g}{2}\right)(-1)^{s_i \cdot x_i} + \frac{\epsilon_g}{2}(-1)^{s_i \cdot \neg x_i}\right]. \end{aligned} \tag{8}$$

Since each $x_i$ and $s_i$ are just binary numbers, we have the following four cases:

- $s_i = 0, x_i = 0$, then $\left(1 - \frac{\epsilon_g}{2}\right)(-1)^{s_i \cdot x_i} + \frac{\epsilon_g}{2}(-1)^{s_i \cdot \neg x_i} = 1$;
- $s_i = 0, x_i = 1$, then $\left(1 - \frac{\epsilon_g}{2}\right)(-1)^{s_i \cdot x_i} + \frac{\epsilon_g}{2}(-1)^{s_i \cdot \neg x_i} = 1$;
- $s_i = 1, x_i = 0$, then $\left(1 - \frac{\epsilon_g}{2}\right)(-1)^{s_i \cdot x_i} + \frac{\epsilon_g}{2}(-1)^{s_i \cdot \neg x_i} = 1 - \epsilon_g$;
- $s_i = 1, x_i = 1$, then $\left(1 - \frac{\epsilon_g}{2}\right)(-1)^{s_i \cdot x_i} + \frac{\epsilon_g}{2}(-1)^{s_i \cdot \neg x_i} = (1 - \epsilon_g) \cdot (-1)$.

It is clear that

$$\left(1 - \frac{\epsilon_g}{2}\right)(-1)^{s_i \cdot x_i} + \frac{\epsilon_g}{2}(-1)^{s_i \cdot \neg x_i} = (1 - \epsilon_g)^{s_i}(-1)^{s_i \cdot x_i}.$$

for all $s_i \in \{0,1\}$ and $x_i \in \{0,1\}$. Thus, based on (8), we have

$$\begin{aligned} E_y[(-1)^{s.x}] &= \prod_{i=1}^{n} \left[\left(1 - \frac{\epsilon_g}{2}\right)(-1)^{s_i \cdot x_i} + \frac{\epsilon_g}{2}(-1)^{s_i \cdot \neg x_i}\right] \\ &= \prod_{i=1}^{n} \left[(1 - \epsilon_g)^{s_i}(-1)^{s_i \cdot x_i}\right] \\ &= (1 - \epsilon_g)^{|s|}(-1)^{s.x}, \end{aligned}$$

where $|s| = \sum_{i=1}^{n} s_i$. Thus, the $\tilde{p}$ with only one bit-flip error is

$$\tilde{p}(x) = E_y[p(y)] = \sum_{s \in \{0,1\}^n} (1 - \epsilon_g)^{|s|} \hat{p}(s)(-1)^{s.x}.$$

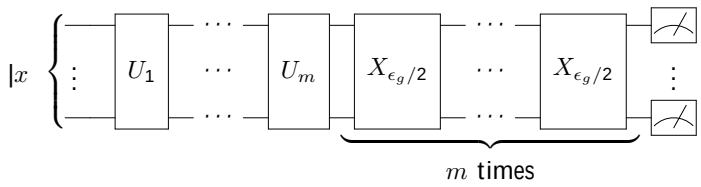



Figure 2: Multiple bit-flip error

### 2.3.2 Extension to Multiple Bit-flip Errors

Now we extend the single bit-flip error model to multiple bit-flip error case, and the circuit we consider is shown in Figure 2. Here, $U_1, ..., U_m$ are noisy-free unitary gates. The resulting distribution in Figure 2 is equivalent to

$$E_{y^{(m)}}[E_{y^{(m-1)}}[...[E_{y^{(1)}}[p(y)]]...]],$$

where $p$ is the error-free output distribution and

$$y_i^{(1)} = \begin{cases} x_i & \text{with probability } 1 - \frac{\epsilon_g}{2} \\ \neg x_i & \text{with probability } \frac{\epsilon_g}{2} \end{cases} \text{ and } y_i^{(j)} = \begin{cases} y_i^{(j-1)} & \text{with probability } 1 - \frac{\epsilon_g}{2} \\ \neg y_i^{(j-1)} & \text{with probability } \frac{\epsilon_g}{2} \end{cases},$$

for $i = 1, ..., n$ and $j = 2, ..., m$. Because the expectations are all over $x$ not $s$, we can repeat the process in Section 2.3.1 $m$ times. Each repetition adds a $(1 - \epsilon_g)^{|s|}$ term in the multiplication:

$$\tilde{p}(x) = E_{y^{(m)}}[E_{y^{(m-1)}}[...[E_{y^{(1)}}[p(y)]]...]] = \sum_{s \in \{0,1\}^n} (1 - \epsilon_g)^{|s| \cdot m} \hat{p}(s)(-1)^{s.x}. \tag{9}$$

However, the circuit in Figure 2 is not valid in practice because an error may occur right after an operator. If all $U_i$ $(1 \le i \le m)$ commute with $X$ up to a global phase factor, then circuit in Figure 2 is equivalent to that in Figure 3, which is realistic in practice.

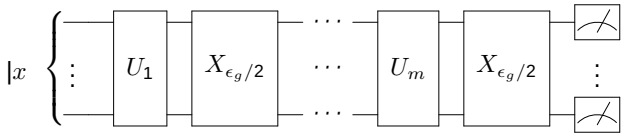



Figure 3: A more practical multiple bit-flip error

### 2.3.3 Bit-flip Error Filter

Let $j_b$ for a $j \in \mathbb{Z}_+$ represents the binary representation of a non-negative integer $j$. Given $\epsilon_g$ and $\tilde{p}(x)$ for all $x \in \{0,1\}^n$, it is possible to recover the noise-free measurement outcomes. The first step is to solve for $\hat{p}(s)$. With known $\epsilon_g$, $\tilde{p}(x)$, and $x$, a linear system derived from Eq. (9) can be built as follows:

$$G\hat{\rho} = \tilde{\rho} \tag{10}$$

where

$$G_{ij} := (1 - \epsilon_g)^{(j-1)m}(-1)^{(i-1)_b \cdot (j-1)_b} \quad \text{for } i \in \{1, ..., 2^n\} \text{ and } j \in \{1, ..., 2^n\}$$

$$\hat{\rho} := \begin{pmatrix} \hat{p}(0...00) \\ \hat{p}(0...01) \\ \vdots \\ \hat{p}(1...11) \end{pmatrix}, \qquad \tilde{\rho} := \begin{pmatrix} \tilde{p}(0...00) \\ \tilde{p}(0...01) \\ \vdots \\ \tilde{p}(1...11) \end{pmatrix},$$

$G \in [-1,1]^{2^n \times 2^n}$, $\hat{\rho} \in \left[-\frac{1}{2^n}, \frac{1}{2^n}\right]^{2^n}$, and $\tilde{\rho} \in [0,1]^{2^n}$. Using the algorithm introduced in Section 3, we can estimate the value of $\epsilon_g$ to construct matrix $G$. Using a sufficient number of measurements, we can compute vector $\tilde{\rho}$. Thus, by solving Eq. (10) and substituting the result into Eq. (7), we can then re-construct the noise-free distribution function $p(x)$ for all $x \in \{0,1\}^n$. The following lemma implies that the solution of Eq. (10) always exists.

**Lemma 2.1.** *$G$ is full-rank for all $n \geq 1$.*

*Proof.* We decompose $G$ as

$$G = G_1^{(n)} \circ G_2^{(n)},$$

where $\circ$ is element-wise multiplication, $G_1^{(n)} \in [0,1]^{2^n \times 2^n}$, $G_2^{(n)} \in \{-1,1\}^{2^n \times 2^n}$, and

$$(G_1^{(n)})_{ij} := (1-\epsilon_g)^{(j-1)m} \quad \text{for } i \in \{1,...,2^n\} \text{ and } j \in \{1,...,2^n\}$$
$$(G_2^{(n)})_{ij} := (-1)^{(i-1)_b (j-1)_b} \quad \text{for } i \in \{1,...,2^n\} \text{ and } j \in \{1,...,2^n\}$$

We start with $G_2^{(n)}$ when $n = 1$. It is easy to examine that

$$G_2^{(1)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

is full-rank. Recall that each entry of $G_2^{(n)}$ is $(-1)^{s.x}$ for binary numbers $s$ and $x$, and each row of $G_2^{(n)}$ shares the same $x$ while each column shares the same $s$. Following the little-endian convention, the 16 entries in $G_2^{(2)}$ can be divided into four divisions equally based on their position

$$\left[\begin{array}{c|c} s_2 = 0, x_2 = 0 & s_2 = 0, x_2 = 1 \\ \hline s_2 = 1, x_2 = 0 & s_2 = 1, x_2 = 1 \end{array}\right].$$

Namely,

$$G_2^{(2)} = \left[\begin{array}{c|c} (-1)^{0\cdot 0} G_2^{(1)} & (-1)^{0\cdot 1} G_2^{(1)} \\ \hline (-1)^{1\cdot 0} G_2^{(1)} & (-1)^{1\cdot 1} G_2^{(1)} \end{array}\right] = \left[\begin{array}{c|c} G_2^{(1)} & G_2^{(1)} \\ \hline G_2^{(1)} & -G_2^{(1)} \end{array}\right].$$

Similarly, we can have

$$G_2^{(n)} = \left[\begin{array}{c|c} G_2^{(n-1)} & G_2^{(n-1)} \\ \hline G_2^{(n-1)} & -G_2^{(n-1)} \end{array}\right] \tag{11}$$

Since $G_2^{(n)} \in \{-1,1\}^{2^n \times 2^n}$, if $G_2^{(n-1)}$ is full-rank, the structure in (11) implies $G_2^{(n)}$ is full-rank. As $G_2^{(1)}$ is full-rank, by induction, $G_2^{(n)}$ is full-rank for all $n \geq 1$.

Because $\epsilon_g \in (0,1)$, $(1-\epsilon_g) = 0$ for all $\epsilon_g \in (0,1)$. Note that for an arbitrary column $j$ of $G$, it is equivalent to column $j$ of $G_2^{(n)}$ multiplied by $(1-\epsilon_g)^{(j-1)m}$. As all columns of $G_2^{(n)}$ are linearly independent, their positive multiples are linearly independent, too. Namely, all columns of $G$ are linearly independent, so $G$ is full-rank for all $n \geq 1$. □

However, similar to the problem in Section 2.2.2, solving Eq. (10) cannot guarantee a meaningful $\tilde{\rho}$, that is, a $\tilde{\rho} \in [0,1]^{2^n}$. Nevertheless, we can consider a optimization problem instead.

$$\begin{aligned} \hat{\rho} := \arg\min \quad & \| G^{-1}\tilde{\rho} - \hat{\rho} \|_2 \\ \text{s.t.} \quad & \\ & \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} \hat{\rho}_i (-1)^{(i-1)_b \cdot (j-1)_b} = 1 \\ & \sum_{i=1}^{2^n} \hat{\rho}_i (-1)^{(i-1)_b \cdot (j-1)_b} \geq 0 \quad \text{for all } j \in \{1,...,2^n\} \end{aligned} \tag{12}$$

As an example, when $n$ = 1, Eq. (7) yields

$$p(0) = \hat{p}(0) + \hat{p}(1)$$
$$p(1) = \hat{p}(0) - \hat{p}(1),$$

so $\hat{p}(0)$ is always $\frac{1}{2}$ as $p(0) + p(1) = 1$. Thus, when $n$ = 1, Eq. (12) can be simplified as

$$\hat{\rho} := \underset{\hat{\rho}_1 = \frac{1}{2}, -\frac{1}{2} \le \hat{\rho}_2 \le \frac{1}{2}}{\arg\min} \quad G^{-1}\tilde{\rho} - \hat{\rho}\ _2 \tag{13}$$

# 3 Estimating Distributions of Noise Parameters

The bit-flip gate error model Eq. (9) and the measurement error model Eq. (2) together provide us forward models to propagate noise in QC algorithms. Based on these forward models and results from a QC device, we can filtering out the bit-flip errors and measurement errors as such to recover noise-free information. Here, a critical step is to identify noise parameters in Eq. (9) and Eq. (2) from multiple measurements of some testing circuits. The Bayesian approach is suited to solving such inverse problems. We will use both the standard Bayesian inference and a novel Bayesian approach [5] to infer these parameters. The key difference of these two approach is that even though both approaches result in a posterior distribution of the desired parameters, the former consider the exact parameters as constants while the latter treat these parameters as random variables intrinsically. In the error model for QC, these two treatments represent different understanding of the underlying physical system (i.e., the device). Namely, the standard Bayesian assumes the exact error rates are deterministic, while the other one assume them to be random variables that may vary.

We take a noisy one-qubit gate $\tilde{U}$ (its error-free version is denoted by $U$) as an example. The probability of the occurrence of a depolarizing error is $\epsilon_g$ as shown in Figure 4. We set the quantity of interest (QoI) in our case to be the probability of measuring 0 from the circuit in Figure 4. Assume

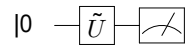



Figure 4: An example testing circuit

that the measurement operator in Figure 4 is associated with measurement errors $\epsilon_{m0}$ and $\epsilon_{m1}$. Let $\lambda := (\epsilon_g, \epsilon_{m0}, \epsilon_{m1})$ be the tuple of noise parameters that we want to infer. Note that if $\tilde{U}$ is a gate like Hadamard gate, the bit-flip error in theory will not affect the measurement outcome for the circuit in Figure 4, which means we only need to infer $\epsilon_{m0}, \epsilon_{m1}$ but not $\epsilon_g$ in this case. Let $\Lambda := (0,1) \times (0,1) \times (0,1)$ denote the space of noise parameters and $D := [0,1]$ denote the space of QoI. Finally, we use $Q : \Lambda \to D$ to denote a general function combining Eq. (9) and Eq. (2) that compute the probability of measuring 0 when testing circuit with bit-flip error and measurement error.

The entire algorithm has two parts. In the first part, $L$ number of QoI's, denoted by $q_j$ for $j = 1, ..., L$, and its corresponding Gaussian kernel density estimation (KDE) are generated from $L$ number of prior $\lambda$'s, denote as $\lambda_j$ for $j = 1, ..., L$, with function $Q$ [5, Algorithm 1]. Then in the second part, prior $\lambda_j$'s are either rejected or accepted based on certain ratio derived from the previous part, and those accepted prior $\lambda_j$'s are the posterior noise parameters that we are looking for [5, Algorithm 2]. Let $\pi_D^{\text{obs}}$ represent the Gaussian KDE of observed probability of measuring 0. The algorithm is listed in Algorithm 1. In practice, the prior $\lambda_j$ are randomly generated from some relatively flat normal distributions due to the little knowledge of its actual characterization.

# 4 Experiments

Because our bit-flip error model solely is not sufficient for the analysis of gate errors in a complicate algorithm like Grover's search or QAOA, the inference for gate errors is done for some simple circuits. For more sophisticated algorithms, we only look at measurement errors. All experiments are

**Algorithm 1** Noise parameters estimation using BJW Bayesian algorithm [5, Algorithm 1&2]

Given a set of prior $\lambda_j$ : $j$ = 1, ..., $L$, Gaussian KDE $\pi_D^{\text{obs}}$ from data, and input state of testing circuit;
**for** $j$ = 1 to $L$ **do**
  Use (9) and (2) with $\lambda_j$ and input state of testing circuit to compute $q_j$;
**end for**
Generate Gaussian KDE $\pi_D^{Q(\text{prior})}$;
Estimate $\mu := \max_\lambda \wedge \frac{\pi_D^{\text{obs}}(Q(\lambda))}{\pi_D^{Q(\text{prior})}(Q(\lambda))}$;
**for** $k$ = 1 to $L$ **do**
  Generate a random number $\zeta_k$ [0, 1] from a uniform distribution;
  Compute ratio $\eta_k := \frac{1}{\mu} \cdot \frac{\pi_D^{\text{obs}}(Q(\lambda_k))}{\pi_D^{Q(\text{prior})}(Q(\lambda_k))}$;
  **if** $\eta_k > \zeta_k$ **then**
    Accept $\lambda_k$;
  **else**
    Reject $\lambda_k$;
  **end if**
**end for**
**output** Accepted noise parameter $\lambda_k$'s.

conducted on IBM's 5-qubit quantum computer `ibmqx2`. Besides comparison between Algorithm 1 and the standard Bayesian method, we also provide comparison to the measurement error filter in Qiskit `CompleteMeasFitter` [2] and the method in [14] based on quantum detector tomography (QDT).

## 4.1 Measurement Errors Filtering Experiment

### 4.1.1 Construction of Error Filter

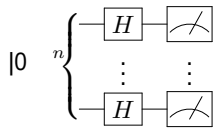
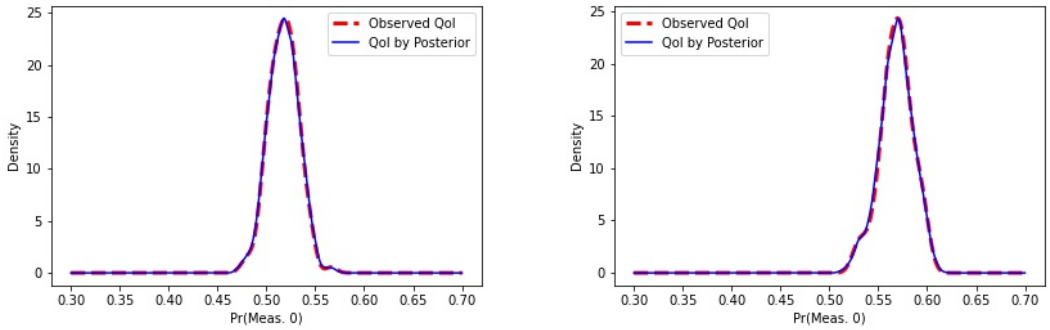



Figure 5: Testing circuit for measurement error parameter inference

In order to infer measurement error parameters in every single qubit, that is, $\epsilon_{m0,i}, \epsilon_{m1,i}$ for $i$ $\{0, 1, 2, 3, 4\}$, in `ibmqx2`, we consider the circuit in Figure 5 where $H$ is the Hadamard gate for each qubit. Theoretically, measurement outcome of $H|0$ is invariant under bit-flip and phase-flip errors, so in this case only measurement error a□ects the distribution of measurement outputs and we do not infer gate error rate $\epsilon_g$. The testing circuit is run for 1024 × 128 times where the fraction of measuring 0 in each ensemble consisting of 1024 runs gives one estimated probability of measuring 0 from the testing circuit. So in total we have 128 data points. For qubit $i$, the prior $(\epsilon_{m0,i}, \epsilon_{m1,i})$ $(0, 1) \times (0, 1)$ are random number from truncated normal distribution $N(\epsilon^0_{m0,i}, 0.1)$ and $N(\epsilon^0_{m1,i}, 0.1)$, respectively, where $\epsilon^0_{m0,i}$ and $\epsilon^0_{m1,i}$ are corresponding values provided by IBM in Qiskit API `IBMQbackend.properties()` after the daily calibration. Then, we use Algorithm 1 to generate the posteriors.

Figure 6 shows that our posteriors approximate the QoI, i.e., the probability of measuring 0, quite well as the almost coincide with the data collected from the device. Tables 1 presents the Kullback–Leibler divergence between two Gaussian KDE for each qubit in Figure 6, which provides a quantitative comparison to illustrate the accuracy of our error model.

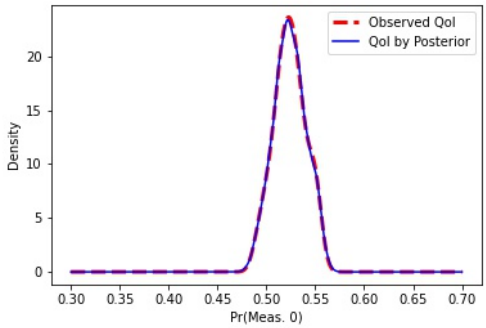
We note that our prior distribution $N(\epsilon^0_{m0,i}, 0.1)$ and $N(\epsilon^0_{m1,i}, 0.1)$ are quite flat and not informative. This is because the provided $\epsilon^0_{m0,i}$ and $\epsilon^0_{m1,i}$ are not always good estimations. If we use relation Eq. (2) and Eq. (6) to construct measurement error filters using the vendor-provided $(\epsilon^0_{m0,i}, \epsilon^0_{m1,i})$ and our posteriors, then apply those filters on the 128 output ensembles of circuit in Figure 5, we obtain di□erent



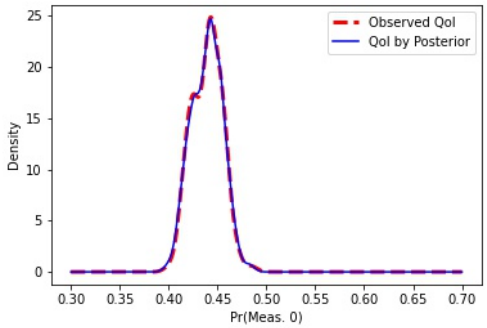

(a) Qubit 1




(b) Qubit 2




(c) Qubit 3




(d) Qubit 4

Figure 6: Gaussian KDE of probability of measuring 0 from data and from values generated by posteriors

results as shown in Figure 7. The ideal probability of measuring 0 for circuit in Figure 5 is clearly 0.5, but the provided parameters rarely gives this value, and its mean and peak of Gaussian KDE are not even close to 0.5. On the other hand, the filters created by our posteriors can make sure mean and peak of the denoised probability of measuring 0 are around the ideal value 0.5.

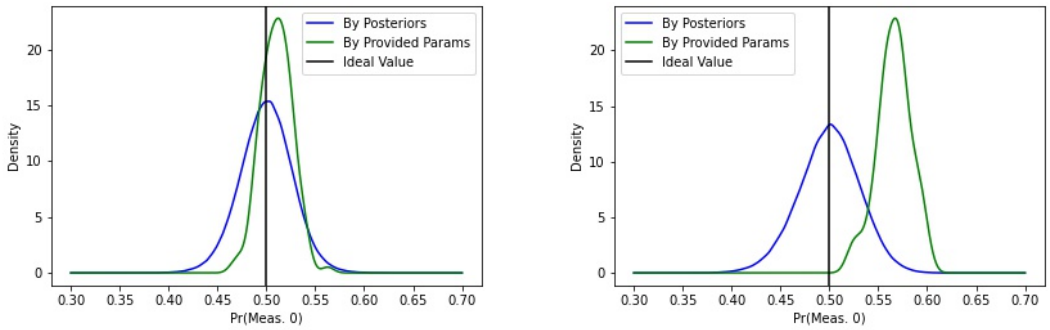
Figure 8 displays the joint and marginal distribution of posteriors. We specifically look at the values of two tuples of posteriors

1. posterior mean: the tuple of posterior $(\epsilon_{m0,i}, \epsilon_{m1,i})$ that are closest to the mean of marginal distributions of posterior $\epsilon_{m0,i}$ and $\epsilon_{m1,i}$ in Euclidean distance;
2. posterior Maximum A Posteriori (MAP): the tuple of posterior $(\epsilon_{m0,i}, \epsilon_{m1,i})$ that are closest to the peaks of marginal distributions of posterior $\epsilon_{m0,i}$ and $\epsilon_{m1,i}$ in Euclidean distance;

The values of posterior mean and posterior MAP are reported in Table 1. We can see in general, $\epsilon_{m1,i}$ is higher than $\epsilon_{m0,i}$, which fits the description in [8].

Table 1: Some statistics from inference

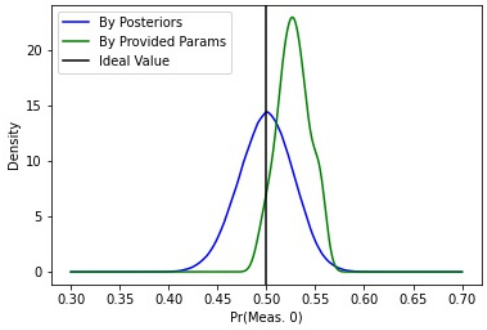
| | Qubit 1 | Qubit 2 | Qubit 3 | Qubit 4 |
|---|---|---|---|---|
| Acceptance Rate | 32.8% | 10.7% | 31.1% | 11.3% |
| KL-div($\pi_p^{Q(\text{posterior})}, \pi_p^{\text{obs}}$) | 0.001014 | 0.002243 | 0.000777 | 0.001610 |
| Post. Mean $(1-\epsilon_{m0,i}, 1-\epsilon_{m1,i})$ | (0.9354,0.9009) | (0.9537,0.8184) | (0.9457,0.8976) | (0.8272,0.9492) |
| Post. MAP $(1-\epsilon_{m0,i}, 1-\epsilon_{m1,i})$ | (0.9797,0.9125) | (0.9863,0.8243) | (0.9858,0.9180) | (0.8426,0.9846) |



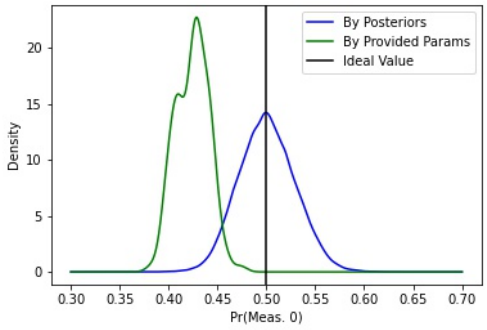

(a) Qubit 1




(b) Qubit 2




(c) Qubit 3




(d) Qubit 4

Figure 7: Gaussian KDE of probability of measuring 0 denoised by provided noise parameters and by posteriors

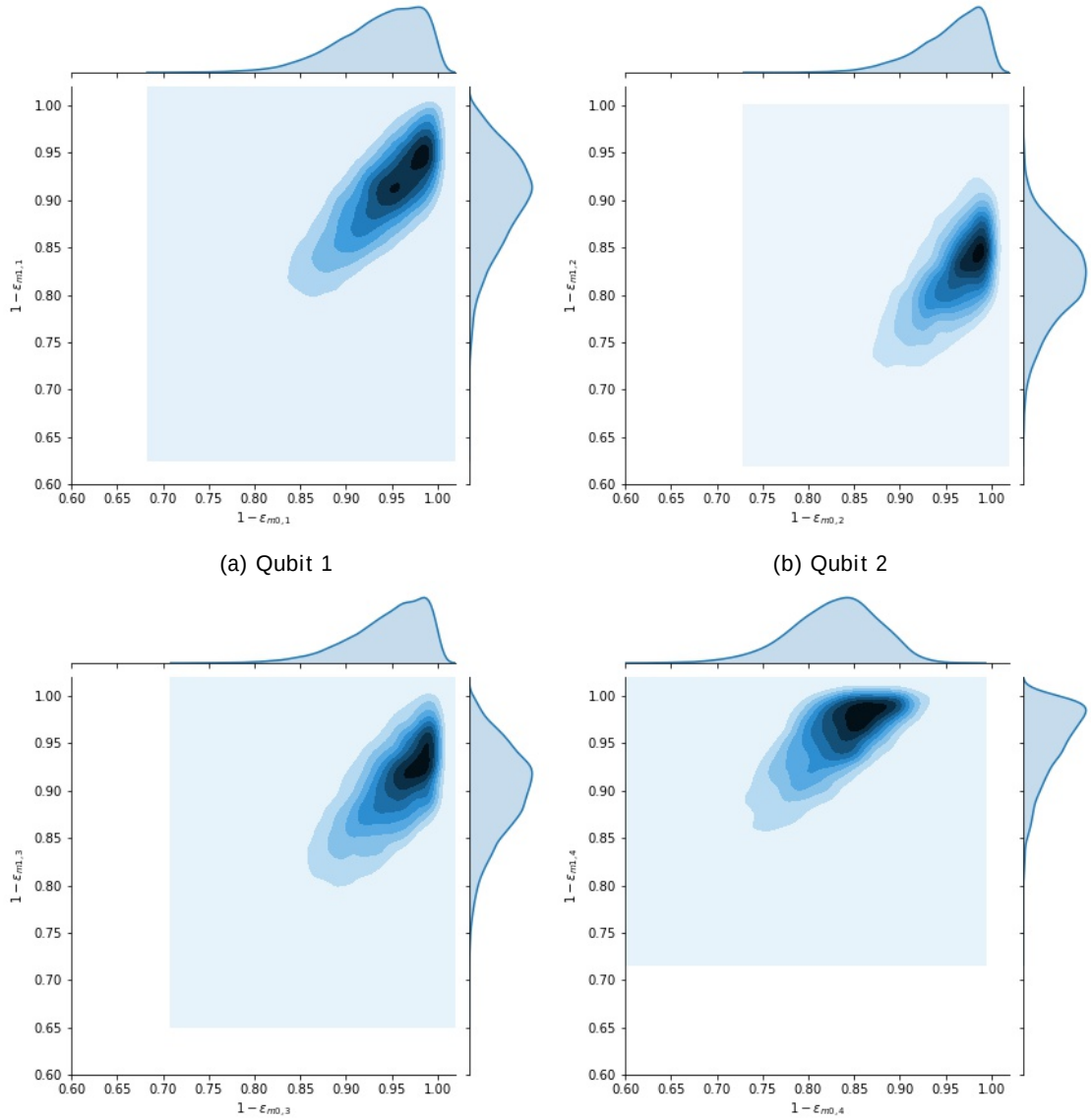
#### 4.1.2 Application on State Tomography

We first apply error filters to the results of state tomography on circuits that make bell basis from |00 and |000 . Qubit 0 and 1 are used for 2-qubit state and Qubit 0 to 2 are used for 3-qubit state in `ibmqx2`. The fidelity between density matrices from (corrected) state tomography result and theoretical quantum state is listed in Table 2. For 2-qubit state tomography, filters constructed from posterior means by Algorithm 1 and standard Bayesian provides similar fidelity as the one by the Qiskit filter. However, for 3-qubit tomography, filters from both Bayesian methods give significantly better fidelity, and their performance are similar. As Qiskit filter assumes correlation in the measurements while our model does not, the fildelity in Table 2 indicates that Bayesian methods provide better choices of parameter values in this experiment.

#### 4.1.3 Application on Grover's Search and QAOA

We apply our filter on Grover's search and QAOA circuits from [1]. We measure Qubit 1 and 2 in `ibmqx2` for Grover's search circuit. The expected solution of this Grover's search example is 11 and the theoretical probability is 1. So in this case, we compare the probability of measuring 11 by running in the real device `ibmqx2` and the denoised probabilities from error filters based on Qiskit method `CompleteMeasFitter`, QDT in [14], mean and MAP of posteriors from standard Bayesian, and mean and MAP of posteriors from Algorithm 1.

As we do not expect the quantum computer has a stable environment, in order to see the robustness of each method in comparison, after the data for creating error filters is collected, we run our Grover's search circuit at several di□erent time and then apply the same set of filters. All results are listed in Table 3.




(a) Qubit 1

(b) Qubit 2

(c) Qubit 3

(d) Qubit 4

Figure 8: Posteriors distributions in the Hadamard gate error test.

As shown in Table 3, both Bayesian methods yield best performance among all the methods while the filters constructed from posterior mean are better than the filters constructed from posterior MAP. In all six time slots, the mean and MAP from Algorithm 1 provide sightly better denoising e□ect than those from standard Bayesian.

The QAOA example has two rounds and parameters for QAOA circuits, as provided in [1], are $(\gamma_1, \beta_1) = (0.2\pi, 0.15\pi)$ and $(\gamma_2, \beta_2) = (0.4\pi, 0.05\pi)$. The graph of the QAOA example in [1] is shown in Figure 9, which has maximum objective value 3 in Max-Cut problem and 6 bit-string optimal solution *l*0010, 0101, 0110, 1001, 1010, 1101 (`ibmqx2` uses little-endien convention, so the rightmost bit is Node 1 and the leftmost bit is Node 4). Because the graph in Figure 9 is a subgraph of the coupling map of `ibmqx2`, we map the nodes to qubits exactly.

The probability of measuring an optimal solution is the quantity to compare in this experiment. Moreover, the results from simulator is also provided as a indicator for the situation without noise. The

Table 2: Fidelity of State Tomography Results Filtered by Various Error Filters

| State | Fidelity | | | |
|---|---|---|---|---|
| | Raw Data | By Qiskit Method | By BJW Mean | By Standard Mean |
| $\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$ | 0.9051 | 0.9800 | 0.9781 | 0.9783 |
| $\frac{1}{\sqrt{2}}(\|01\rangle + \|10\rangle)$ | 0.9157 | 0.9803 | 0.9806 | 0.9808 |
| $\frac{1}{\sqrt{2}}(\|000\rangle + \|111\rangle)$ | 0.7389 | 0.9227 | 0.9390 | 0.9391 |
| $\frac{1}{\sqrt{2}}(\|010\rangle + \|101\rangle)$ | 0.6719 | 0.8970 | 0.9203 | 0.9207 |
| $\frac{1}{\sqrt{2}}(\|100\rangle + \|011\rangle)$ | 0.7006 | 0.9121 | 0.9254 | 0.9207 |
| $\frac{1}{\sqrt{2}}(\|110\rangle + \|001\rangle)$ | 0.6974 | 0.8863 | 0.9443 | 0.9446 |

* "Qiskit Method" means to `CompleteMeasFitter` in Qiskit [2]. "BJW mean" implies the transition matrix is created from posterior mean by Algorithm 1. "Standard Mean" means the transition matrix is created from posterior mean by standard Bayesian

Table 3: Probability of measuring 11 in Grover's search example

| Method/Source | Hour 0 | Hour 2 | Hour 4 | Hour 8 | Hour 12 | Hour 16 |
|---|---|---|---|---|---|---|
| Raw Data | 0.6727 | 0.6930 | 0.6724 | 0.6740 | 0.6917 | 0.6841 |
| Qiskit Method | 0.7097 | 0.7335 | 0.7104 | 0.7120 | 0.7323 | 0.7241 |
| QDT | 0.7107 | 0.7332 | 0.7087 | 0.7108 | 0.7305 | 0.7224 |
| Standard Mean | 0.9099 | 0.9324 | 0.9063 | 0.9088 | 0.9290 | 0.9192 |
| Standard MAP | 0.8378 | 0.8635 | 0.8372 | 0.8392 | 0.8616 | 0.8522 |
| BJW Mean | 0.9128 | 0.9351 | 0.9088 | 0.9114 | 0.9316 | 0.9219 |
| BJW MAP | 0.8920 | 0.9158 | 0.8914 | 0.8936 | 0.9128 | 0.9034 |

* "Qiskit Method" means to `CompleteMeasFitter` in Qiskit [2], QDT refers to filer in [14], "Standard" stands for Standard Bayesian, and BJW is Algorithm 1. MAP and mean represent the error filters are created from the MAP and mean of posteriors.
** "Hour X" means the experiment is conducted X hours after the data for error filers of all listed methods are collected

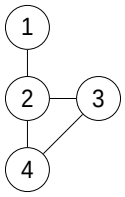



Figure 9: The graph of QAOA example in [1]

remaining procedures are the same as those in Grover's search experiment. The data is reported in Table 4. In this part of the experiment, "QDT" error filter from [14] is built under the assumption that measurement operations are independent between each qubit due to the large amount of testing circuits that correlation assumption requires ($6^4$ circuits if assume qubits are correlated in measurement) .

Table 4: Probability of measuring an optimal solution in QAOA example

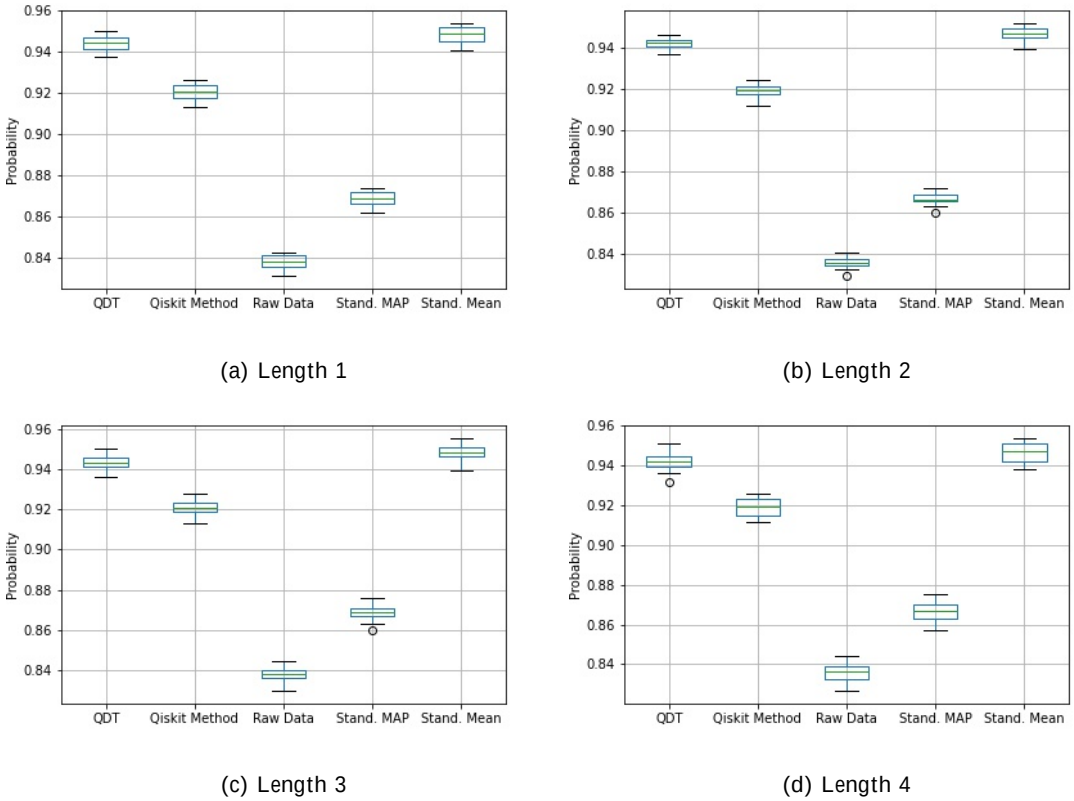
| Method/Source | Hour 0 | Hour 2 | Hour 4 | Hour 8 | Hour 12 | Hour 16 |
|---|---|---|---|---|---|---|
| Simulator | 0.8930 | 0.8937 | 0.8941 | 0.8943 | 0.8940 | 0.8940 |
| Raw Data | 0.5784 | 0.6038 | 0.5895 | 0.5725 | 0.5748 | 0.5740 |
| Qiskit Method | 0.5968 | 0.6456 | 0.6316 | 0.6074 | 0.6140 | 0.6155 |
| QDT | 0.6400 | 0.6698 | 0.6525 | 0.6312 | 0.6331 | 0.6325 |
| Standard Mean | 0.6952 | 0.7239 | 0.7033 | 0.6766 | 0.6787 | 0.6797 |
| Standard MAP | 0.6444 | 0.6695 | 0.6508 | 0.6305 | 0.6309 | 0.6317 |
| BJW Mean | 0.6975 | 0.7265 | 0.7058 | 0.6790 | 0.6810 | 0.6822 |
| BJW MAP | 0.6610 | 0.6860 | 0.6672 | 0.6431 | 0.6439 | 0.6452 |

See notes in Table 3 for the explanation of the column and row names.

The conclusion derived from Table 4 is basically the same as the conclusion from Table 3. Namely, Bayesian methods, especially filters from posterior mean, outperform other methods, and parameters inferred by Algorithm 1 works slightly better than those by standard Bayesian in all six time slots. From both Grover's search and QAOA examples, we can see that treating noise parameters as random variables is more suited to modeling the error rates.

#### 4.1.4 Application on Random Clifford Circuits

We also test the measurement-error filtering for random 2-Qubit Cli□ord circuits with 1, 2, 3, and 4 2-Qubit Cli□ord operators (i.e., length 1, 2, 3, 4). For each length, 16 random circuits are generated to draw a boxplot and each circuit is run for 8192 shots. The results are shown in Figure 10. While the theoretical output of 2-Qubit Cli□ord circuit is 00 with probability 1, Figure 10 demonstrate that the filter constructed from posterior mean estimated by standard Bayesian provides best performance. Algorithm 1 results in almost the same results as the standard Bayesian method.




(a) Length 1

(b) Length 2

(c) Length 3

(d) Length 4

Figure 10: Measurement-error filtering for random 2-Qubit Cli□ord circuits. "Length" represents the number of Cli□ord operators in the circuit. "Probability" means the probability of measuring 00.

## 4.2 Gate and Measurement Error Filtering Experiment

simple circuit, We consider the circuit with 200 NOT gates as shown in Figure 11. We still use machine `ibmqx2` and run the experiment twice separately on Qubit 1 and Qubit 2. In each trial, the circuit is run $1024 \times 128$ times where readouts from every 1024 runs are used to estimate the probability of measuring 0. Note that the gate error model (9) require us to know the noise-free probability distribution of all basis if we want to infer $\epsilon_g$. So we use experiment in two ways: assume we know noise-free probability distribution then infer $(\epsilon_{m0}, \epsilon_{m1}, \epsilon_g)$ and assume we know $(\epsilon_{m0}, \epsilon_{m1}, \epsilon_g)$ and do denoising.

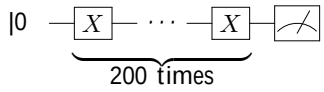
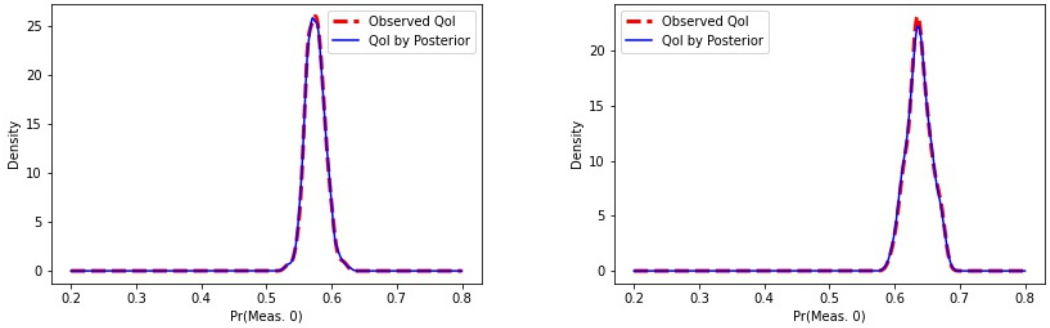



Figure 11: Experiment Circuit

Because the previously mentioned Qiskit method `CompleteMeasFitter` and QDT are for measurement errors, in this section, we only compare the results from standard Bayesian and Algorithm 1 with same priors and dataset. Note that our QoI is still the probability of measuring 0 when we use Algorithm 1 to create filters. The priors are truncated normal $N(\epsilon^0_{m0,i}, 0.1)$, $N(\epsilon^0_{m1,i}, 0.1)$, and $N(\epsilon^0_g, 0.01)$ with range $(0, 1)$. Again, $\epsilon^0_{m0,i}, \epsilon^0_{m1,i}, \epsilon^0_g$ are provided values from IBM's daily calibration and standard deviations are chosen to make prior relatively flat due to the lack of knowledge on these parameters.

### 4.2.1 Inference for Noise Parameters

Figure 12 illustrates that Algorithm 1 still produces posteriors that results in a very nice match on our data. Figure 13 shows the distribution of $\epsilon_g$ in Qubit 1 and 2. Both distributions are right-skewed. Table 5 provides the numerical values for mean and MAP. In Table 5, we can see both methods give similar measurement error parameter $\epsilon_{m0,i}$ and $\epsilon_{m1,i}$ on Qubit 1 and 2, but the gate error rate $\epsilon_g$ are not always similar. However, as shown in Figure 14, posteriors from Algorithm 1 can generate QoI that fits the entire distribution of data while the QoI generated by posteriors from standard Bayesian can only match data in mean or MAP (as our distribution of data is quite symmetric).




(a) Qubit 1

(b) Qubit 2

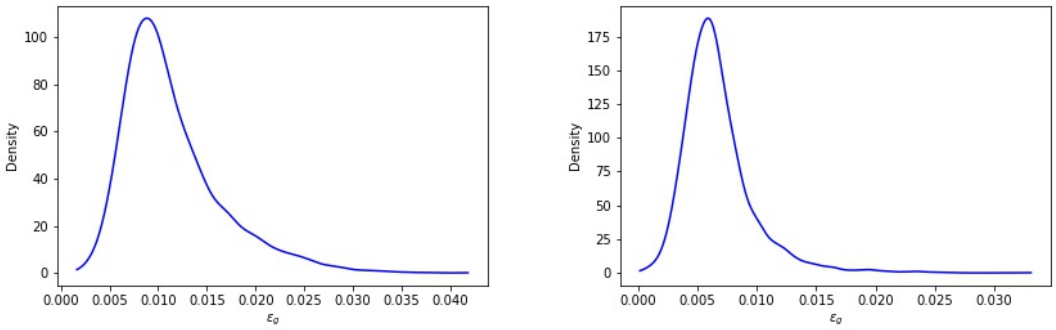
Figure 12: Gaussian KDE of probability of measuring 0 from data and from values generated by posteriors for circuit in Figure 11.

Table 5: Some statistics from inference by both Bayesian methods

| | Qubit 1 | Qubit 2 |
|---|---|---|
| BJW Acceptance Rate | 15.6% | 11.4% |
| BJW KL-div($\pi_D^{Q(\text{posterior})}, \pi_D^{\text{obs}}$) | 0.001381 | 0.000844 |
| BJW Post. Mean $(1-\epsilon_{m0,i}, 1-\epsilon_{m1,i}, \epsilon_g)$ | (0.9255, 0.8922, 0.009868) | (0.9229, 0.8856, 0.007609) |
| BJW Post. MAP $(1-\epsilon_{m0,i}, 1-\epsilon_{m1,i}, \epsilon_g)$ | (0.9756, 0.8837, 0.009654) | (0.9770, 0.9485, 0.006582) |
| Standard Post. Mean $(1-\epsilon_{m0,i}, 1-\epsilon_{m1,i}, \epsilon_g)$ | (0.9221, 0.8939, 0.009367) | (0.9214, 0.8871, 0.005965) |
| Standard Post. MAP $(1-\epsilon_{m0,i}, 1-\epsilon_{m1,i}, \epsilon_g)$ | (0.9758, 0.8835, 0.013101) | (0.9836, 0.9354, 0.006906) |

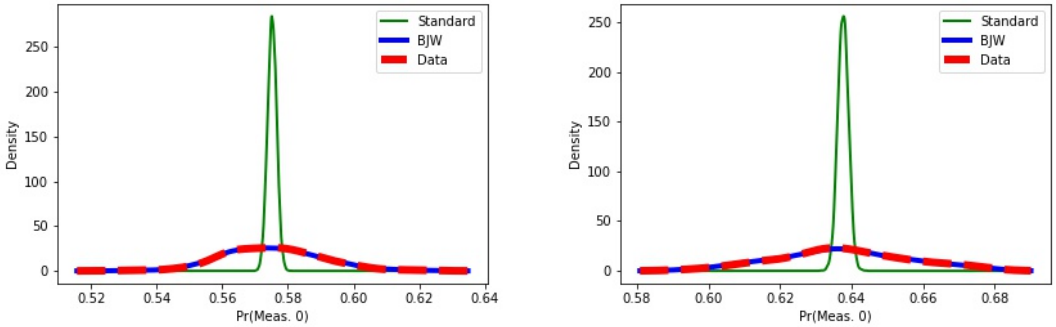
### 4.2.2 Error Filtering

Using the posterior means from Table 5, we construct gate and measurement error filters and apply them on our 128 probabilities of measuring 0 on Qubit 1 and Qubit 2. The results are displayed in Figure 15.




(a) Qubit 1 (b) Qubit 2

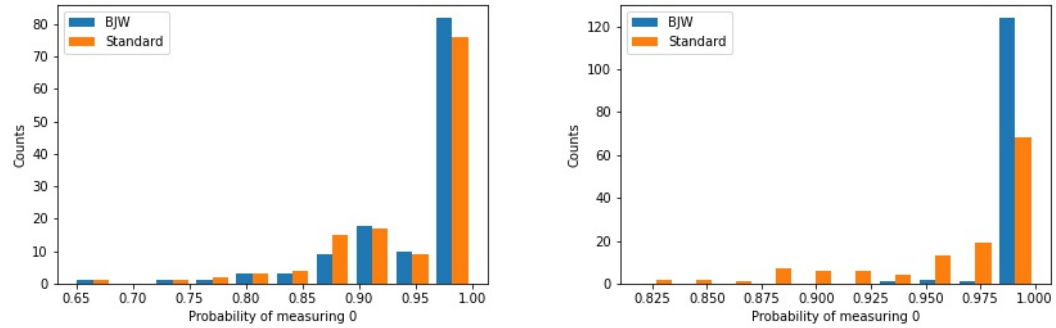
Figure 13: Gaussian KDE of $\epsilon_g$




(a) Qubit 1 (b) Qubit 2

Figure 14: Gaussian KDE of QoI generated by posteriors from both Bayesian methods

In the most of time, the filter created by parameters from Algorithm 1 can recover the ideal probability 1. Again, we can also see filters from Algorithm 1 are clearly better those from standard Bayesian, especially for Qubit 2.




(a) Qubit 1 (b) Qubit 2

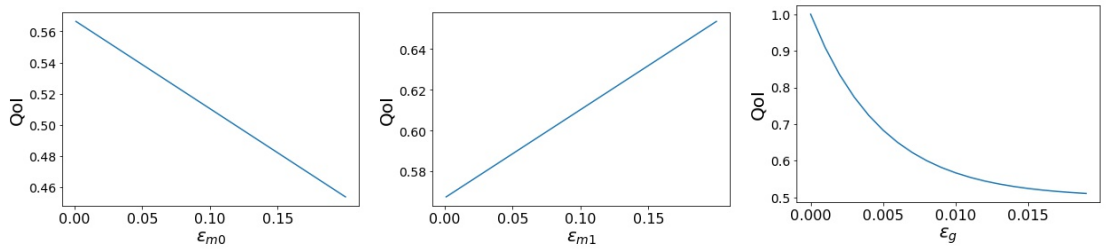
Figure 15: Denoised (both gate and measurement) probability of measuring 0. Parameters used are posterior mean from Table 5.

The data for error filters in Section 4.1 and experiment in Section 4.2 were collected within one hour, so it is reasonable to use posteriors in Section 4.2 to denoise the Grover's search data in Section 4.1. However, comparing the values of measurement error parameters in Table 1 and in Table 5, we can see there are some noticeable di□erences. Table 6 gives the performance of using parameters in Section 4.2 to filter out errors in data in Section 4.1. Comparing those values to the ones in Table 3, we can see they are better than Qiskit method and QDT, but worse than values from either Bayesian methods.

Table 6: Probability of measuring 11 in Grover's search Example denoised by parameters in Section 4.2

| Method | Hour 0 | Hour 2 | Hour 4 | Hour 8 | Hour 12 | Hour 16 |
|---|---|---|---|---|---|---|
| Standard Mean | 0.8398 | 0.8680 | 0.8392 | 0.8414 | 0.8656 | 0.8561 |
| Standard MAP | 0.8116 | 0.8367 | 0.8110 | 0.8131 | 0.8348 | 0.8257 |
| BJW Mean | 0.8434 | 0.8716 | 0.8428 | 0.8450 | 0.8691 | 0.8595 |
| BJW MAP | 0.7992 | 0.8240 | 0.7986 | 0.8005 | 0.8221 | 0.8131 |

One possible explanation is, with 200 gates, our model is much more sensitive to gate error than measurement errors, as Figure 16 demonstrates. In this case, $\epsilon_g$ is estimated more precisely than $\epsilon_{m0}, \epsilon_{m1}$ during the experiment in Section 4.2. On the other hand, the estimation of measurement error in Section 4.1 uses circuits that have 0.5 chance to measuring either 0 or 1 without noise and this distribution does not change under the depolarizing error assumption on Hadamard gate. In this case, $\epsilon_{m0}$ and $\epsilon_{m1}$ should be estimated equal-precisely in Section 4.1 and thus provide a better performance than parameters from Section 4.2.




(a) $\epsilon_{m1} = 0, \epsilon_g = 0.01$ (b) $\epsilon_{m0} = 0, \epsilon_g = 0.01$ (c) $\epsilon_{m1} = 0, \epsilon_{m0} = 0$

Figure 16: Sensitivity analysis of forward model $Q$ with 200 NOT gate.

# 5 Discussion and Future Works

In this work, we extend a bit-flip error model from single gate case to multiple gate case, and provide theoretical analysis to prove the existence of the error mitigation solution for both cases. We propose to use Bayesian approaches to infer parameters in the error models, so as to characterize the propagation of the device noise in QC algorithms more e□ectively. The experiments in Section 4 demonstrate that our methodology outperforms two existing methods on the same error models over a wide range of time, while the number of testing circuits is linear or constant to the number of qubits. The modified Bayesian approach is, in general, better than the standard Bayesian. These results indicates that our error models can characterize the device noise quite well, and they help to understand the propagation of such noise in QC algorithms. Consequently, the error mitigation approaches based on these error models yield good results in our tests.

There are still several limitation in our methodology. One issue that a□ects the scalability of our method is the exponentially large matrix in the denoising step. The dimension of matrix can be reduced if we can identity the qubits that are independent during the measurement step of an algorithm and filter their measurement outcome separately. On the other hand, because the method of estimating the distribution of model parameters is not limited to the two models we discussed in the paper, a new measurement error model discussed in [4] probably provide a way around for correlated measurement error and exponentially large transition matrix. Also, our gate error model can only deal with bit-flip

error in some simple circuits. More effects are required to further improve our error mitigation model as such to make it more general and accurate.

## 6 Acknowledgement

This work is supported by Defense Advanced Research Projects Agency as part of "The Quantum Computing Revolution and Optimization: Challenges and Opportunities" project.

## References

[1] J. Abhijith, Adetokunbo Adedoyin, John Ambrosiano, Petr Anisimov, Andreas Bärtschi, William Casper, Gopinath Chennupati, Carleton Coffrin, Hristo Djidjev, David Gunter, Satish Karra, Nathan Lemons, Shizeng Lin, Alexander Malyzhenkov, David Mascarenas, Susan Mniszewski, Balu Nadiga, Daniel O'Malley, Diane Oyen, Scott Pakin, Lakshman Prasad, Randy Roberts, Phillip Romero, Nandakishore Santhi, Nikolai Sinitsyn, Pieter J. Swart, James G. Wendelberger, Boram Yoon, Richard Zamora, Wei Zhu, Stephan Eidenbenz, Patrick J. Coles, Marc Vuffray, and Andrey Y. Lokhov. Quantum Algorithm Implementations for Beginners. *arXiv e-prints*, page arXiv:1804.03719, April 2018.

[2] Gadi Aleksandrowicz, Thomas Alexander, Panagiotis Barkoutsos, Luciano Bello, Yael Ben-Haim, David Bucher, Francisco Jose Cabrera-Hernández, Jorge Carballo-Franquis, Adrian Chen, Chun-Fu Chen, Jerry M. Chow, Antonio D. Córcoles-Gonzales, Abigail J. Cross, Andrew Cross, Juan Cruz-Benito, Chris Culver, Salvador De La Puente González, Enrique De La Torre, Delton Ding, Eugene Dumitrescu, Ivan Duran, Pieter Eendebak, Mark Everitt, Ismael Faro Sertage, Albert Frisch, Andreas Fuhrer, Jay Gambetta, Borja Godoy Gago, Juan Gomez-Mosquera, Donny Greenberg, Ikko Hamamura, Vojtech Havlicek, Joe Hellmers, Łukasz Herok, Hiroshi Horii, Shaohan Hu, Takashi Imamichi, Toshinari Itoko, Ali Javadi-Abhari, Naoki Kanazawa, Anton Karazeev, Kevin Krsulich, Peng Liu, Yang Luh, Yunho Maeng, Manoel Marques, Francisco Jose Martín-Fernández, Douglas T. McClure, David McKay, Srujan Meesala, Antonio Mezzacapo, Nikolaj Moll, Diego Moreda Rodríguez, Giacomo Nannicini, Paul Nation, Pauline Ollitrault, Lee James O'Riordan, Hanhee Paik, Jesús Pérez, Anna Phan, Marco Pistoia, Viktor Prutyanov, Max Reuter, Julia Rice, Abdón Rodríguez Davila, Raymond Harry Putra Rudy, Mingi Ryu, Ninad Sathaye, Chris Schnabel, Eddie Schoute, Kanav Setia, Yunong Shi, Adenilton Silva, Yukio Siraichi, Seyon Sivarajah, John A. Smolin, Mathias Soeken, Hitomi Takahashi, Ivano Tavernelli, Charles Taylor, Pete Taylour, Kenso Trabing, Matthew Treinish, Wes Turner, Desiree Vogt-Lee, Christophe Vuillot, Jonathan A. Wildstrom, Jessica Wilson, Erick Winston, Christopher Wood, Stephen Wood, Stefan Wörner, Ismail Yunus Akhalwaya, and Christa Zoufal. Qiskit: An Open-source Framework for Quantum Computing, January 2019.

[3] H. Bombin, Ruben S. Andrist, Masayuki Ohzeki, Helmut G. Katzgraber, and M. A. Martin-Delgado. Strong resilience of topological codes to depolarization. *Physical Review X*, 2(2), April 2012.

[4] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. Mckay, and Jay M. Gambetta. Mitigating measurement errors in multi-qubit experiments. *arXiv e-prints*, page arXiv:2006.14044, June 2020.

[5] T. Butler, J. Jakeman, and T. Wildey. Combining push-forward measures and bayes' rule to construct consistent solutions to stochastic inverse problems. *SIAM Journal on Scientific Computing*, 40(2):A984–A1011, January 2018.

[6] Yanzhu Chen, Maziar Farahzad, Shinjae Yoo, and Tzu-Chieh Wei. Detector tomography on IBM quantum computers and mitigation of an imperfect measurement. *Physical Review A*, 100(5), November 2019.

[7] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Physical Review X*, 8(3), July 2018.

[8] IBM Quantum Experience. Backend properties.

[9] Jaromír Fiurášek. Maximum-likelihood estimation of quantum measurement. *Physical Review A*, 64(2), July 2001.

[10] Michael R Geller. Rigorous measurement error correction. *Quantum Science and Technology*, 5(3):03LT01, June 2020.

[11] 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(7749):491–495, 2019.

[12] Aleksander Marek Kubica. *The ABCs of the Color Code: A Study of Topological Quantum Codes as Toy Models for Fault-Tolerant Quantum Computation and Quantum Phases Of Matter*. PhD thesis, 2018.

[13] Raymond Laflamme, Cesar Miquel, Juan Pablo Paz, and Wojciech Hubert Zurek. Perfect quantum error correcting code. *Physical Review Letters*, 77(1):198–201, July 1996.

[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, April 2020.

[15] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2010.

[16] Ryan ODonnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[17] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018.

[18] Yasuhiro Takahashi, Yuki Takeuchi, and Seiichiro Tani. Classically Simulating Quantum Circuits with Local Depolarizing Noise. *arXiv e-prints*, page arXiv:2001.08373, January 2020.

[19] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, 119(18), November 2017.

[20] Colin J Trout, Muyuan Li, Mauricio Gutiérrez, Yukai Wu, Sheng-Tao Wang, Luming Duan, and Kenneth R Brown. Simulating the performance of a distance-3 surface code in a linear ion trap. *New Journal of Physics*, 20(4):043038, April 2018.