

# Optimizing Circuit Reusing and its Application in Randomized Benchmarking

Zhuo Chen, Guoding Liu, and Xiongfeng Ma

Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084 China

Quantum learning tasks often leverage randomly sampled quantum circuits to characterize unknown systems. An efficient approach known as “circuit reusing,” where each circuit is executed multiple times, reduces the cost compared to implementing new circuits. This work investigates the optimal reusing times that minimizes the variance of measurement outcomes for a given experimental cost. We establish a theoretical framework connecting the variance of experimental estimators with the reusing times  $R$ . An optimal  $R$  is derived when the implemented circuits and their noise characteristics are known. Additionally, we introduce a near-optimal reusing strategy that is applicable even without prior knowledge of circuits or noise, achieving variances close to the theoretical minimum. To validate our framework, we apply it to randomized benchmarking and analyze the optimal  $R$  for various typical noise channels. We further conduct experiments on a superconducting platform, revealing a non-linear relationship between  $R$  and the cost, contradicting previous assumptions in the literature. Our theoretical framework successfully incorporates this non-linearity and accurately predicts the experimentally observed optimal  $R$ . These findings underscore the broad applicability of our approach to experimental realizations of quantum learning protocols.

## 1 Introduction

The development of large-scale and functional quantum computers requires the efficient characterization of quantum systems, a task often referred to as “quantum learning” [1–3]. This includes estimating observables of quantum states [4–6] and assessing noise in quantum channels [7–9]. A wealth of techniques within quantum learning have been proposed with the aim of efficiently evaluating the properties of quantum states and channels. Among these, the “randomization” approach is pivotal, where outcomes from random circuits are averaged to derive desired quantities. This approach involves sampling random quantum states [10], employing random gate sequences [11–13], and implementing randomized measurements [14–16].

Within the context of quantum learning tasks, minimizing resource consumption, such as the number of implemented circuits and experimental duration, is crucial. A viable

Zhuo Chen: These authors contributed equally to this work.

Guoding Liu: These authors contributed equally to this work.

Xiongfeng Ma: xma@tsinghua.edu.cn

strategy to achieve this is “circuit reusing”, where each randomly sampled circuit is used multiple times, including state preparation, gate implementation, and measurement. This strategy is extensively applied in the experimental execution of quantum learning tasks, such as classical shadow [17] and randomized benchmarking [18]. In experiments, initializing a new random circuit generally consumes more time than reusing an existing one due to necessary steps like random sampling, quantum gate decomposition, and other hardware preparations. Therefore, employing a circuit reusing strategy can reduce experimental costs while maintaining a reasonable data sample volume.

Theoretically, increasing the number of different circuits can enhance randomness and, thereby, improve estimation precision [3]. This introduces a trade-off between cost savings and improved estimation accuracy. An important question emerges: what is the optimal number of times a circuit should be reused to minimize the estimation variance for a given experimental cost? This issue has been explored in the context of classical shadows [19, 20], analyzed in a noiseless scenario, with a cost model suggesting that executing a circuit  $R$  times incurs a cost  $t(R) = R + \alpha$ , where  $\alpha$  represents the fixed overhead associated with initializing a new circuit.

When considering the reusing strategy in broader quantum learning tasks and practical scenarios, several challenges arise. Firstly, unavoidable circuit noise must be considered in the assessment of the optimal reusing times, particularly for tasks aimed at noise benchmarking [12, 21]. Furthermore, experimental data indicates that the assumption of a constant overhead  $\alpha$  is unrealistic, necessitating a more universal cost model applicable to diverse quantum learning tasks.

In this study, we examine how the number of reusing times,  $R$ , influences the variance of outcomes in diverse quantum learning tasks. We introduce a cost model,  $T(N, R) = Nt(R)$ , to describe the cost of implementing  $N$  distinct circuits, each reused  $R$  times, where  $t(R)$  varies as a function of  $R$ . With this model, we formulate an optimization issue that aims to minimize the variance of the target estimator as a function of the reusing times  $R$  for a fixed experimental budget.

Employing the Law of Total Variance, we decompose the variance of the estimator into two components and establish the relationship between these components and the reusing times  $R$ . This analysis enables us to identify the optimal  $R$  that minimizes variance for various cost models  $t(R)$  and across different quantum learning tasks. While the value of the optimal reusing times depends on the experimental configuration, implemented circuits, and their specific noise channel characteristics, we introduce a practical “near-optimal” solution that requires no prior knowledge of the noise channel or the executed circuits. We prove that the variance of the near-optimal solution does not exceed a constant multiple of the minimal achievable variance given a bound on  $t(R)$ . Notably, both the optimal and the near-optimal strategies are versatile and applicable to various quantum learning tasks such as randomized benchmarking (RB) [11, 12, 21], shadow estimation [13], direct fidelity estimation [22], and quantum tomography [23, 24]. Additionally, the principles underlying this solution can enhance other techniques that utilize the “circuit reusing” strategy, including quantum error mitigation [25] and entanglement detection [26].

As an application of our findings, we explore the reusing strategy within the standard RB protocol [21] and identify the optimal reusing times  $R^*$  with numerical simulations across typical noise channels. Moreover, we conduct 2-qubit standard RB experiments with different reusing times  $R$  ranging from 2 to 20000 on a superconducting quantum computing platform [27]. Our experimental findings reveal that the simplest linear cost model  $t(R) = R + \alpha$  does not hold. This observation leads us to develop a tailored cost model specific to our experimental setup. In experiments, the optimal reusing times  $R^*$

ranges from 100 to 500, aligning closely with our theoretically predicted optimal solution of 200. Furthermore, our near-optimal solution, with  $R_0 = 333$ , exhibits a variance below 2 times the minimal variance observed experimentally, confirming our theoretical predictions.

The structure of this paper is organized as follows. Section 2 introduces our analytical framework and explores the optimization of reusing times  $R$  under various experimental conditions. Section 3 applies our findings to RB and presents corresponding numerical simulations across various noise channels. Section 4 details the experimental demonstration conducted on a superconducting quantum computing platform. Finally, we conclude in Section 5 with a summary of our results and implications for future research.

## 2 Optimize circuit reusing in quantum learning tasks

### 2.1 Formulation of the optimization issue for circuit reusing times

The task of quantum learning aims to assess the property of quantum states or quantum channels via measurement results from multiple circuits. When the protocol of the target task involves one-shot measurement across distinct circuits, it is feasible to perform repeated measurements to reuse circuits. Below, we analyze the relationship between the estimation variance and the reusing times.

As depicted in Fig. 1, let us outline the steps involved in a quantum learning task: We execute  $N$  random circuits, with each circuit potentially involving distinct state  $\rho_i$ , gate sequence  $\vec{G}_i$ , and measurement  $Q_i$ . Implementation of each circuit is repeated  $R$  times. For the  $i$ -th circuit, the  $r$ -th single-shot measurement result is represented as a random variable, denoted by  $X_r(\rho_i, Q_i, \vec{G}_i)$ . The target estimator  $\bar{X}_N(R)$  is then defined as

$$\begin{aligned}\bar{X}_N(R) &= \frac{1}{N} \sum_i \bar{X}_R(\rho_i, Q_i, \vec{G}_i), \\ \bar{X}_R(\rho_i, Q_i, \vec{G}_i) &= \frac{1}{R} \sum_r X_r(\rho_i, Q_i, \vec{G}_i).\end{aligned}\tag{1}$$

More generally,  $X_r(\rho_i, Q_i, \vec{G}_i)$  can be a value of the  $r$ -th single-shot measurement result after classical post-processing. For our further derivation, we only require the independent and identically distributed condition for  $X_r(\rho_i, Q_i, \vec{G}_i)$  over different  $r$  and for  $\bar{X}_R(\rho_i, Q_i, \vec{G}_i)$  over different random gate sequences. This is ensured by the uniform and random sampling for gate sequences and the independence among different single-shot measurement results.

To identify the optimal value of  $R$  to minimize the variance of  $\bar{X}_N(R)$ , we develop the subsequent optimization model. Suppose implementing each circuit  $R$  times consumes cost  $t(R)$ , which typically pertains to time expense, the total cost  $T_0$  for acquiring  $NR$  samples can be described as

$$T_0 = Nt(R).\tag{2}$$

The situation can further be generalized to the case that  $t(R)$  also depends on  $N$ . With this foundation, we formulate the optimization question as follows.

**Question 1.** *Given a predetermined total cost  $T_0$ , the objective is to determine the optimal reusing times  $R$  that minimizes the variance of the estimator  $\bar{X}_N$ . Mathematically,*

$$\begin{aligned}\arg \min_R \mathbb{V}(\bar{X}_N) \\ s.t. \quad Nt(R) = T_0.\end{aligned}\tag{3}$$

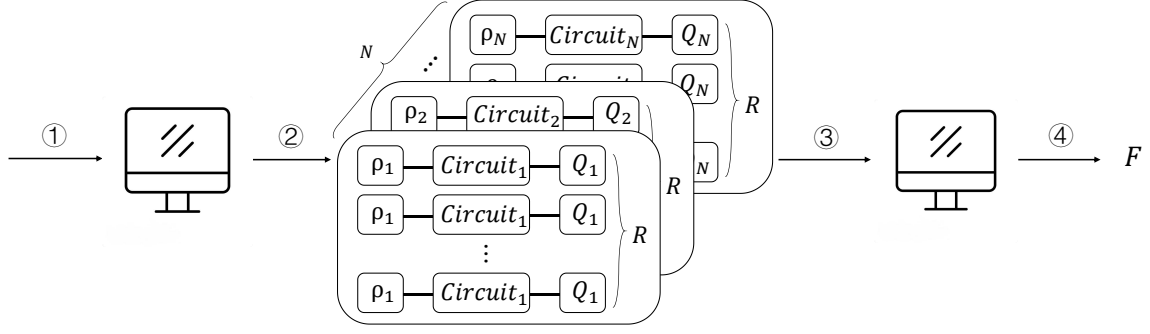


Figure 1: Procedure of a quantum learning task utilizing circuit reusing: ①Input the quantum learning objective and the prior knowledge about quantum device performance and implemented circuits to a classical computer for analysis. ②Design the learning protocol, setting parameters such as the reusing times  $R$  and the number of unique circuits  $N$ , and generate the corresponding quantum circuits. ③Execute quantum operations and measurements to gather data for estimating the target quantity. ④Analyze the collected data using classical post-processing to derive the desired quantity.

In previous works [19, 20], an optimization issue was analyzed for the strategy of circuit reusing within the framework of randomized shadow estimation, assuming a linear cost model of  $t(R) = R + \alpha$ , where  $\alpha$  is a positive constant representing the extra cost incurred when initializing a new circuit compared to reusing an existing one. In contrast, our approach accommodates more general scenarios. Observations from RB experiments suggest that the linear model may not be applicable. This variability is substantiated by runtime data from our RB experiments conducted on a superconducting quantum computing platform, which we detail in Section 4. Consequently, our optimization issue, as outlined in Question 1, requires a more universal expression for  $t(R)$ .

Assuming that the circuits  $\{\rho_i, Q_i, \vec{G}_i\}$  are uniformly and independently determined, the variance of the estimator is given by:

$$\mathbb{V}(\bar{X}_N(R)) = \frac{1}{N} \mathbb{V}(\bar{X}_R(\rho_i, Q_i, \vec{G}_i)), \quad (4)$$

where  $\bar{X}_R$  is the average over  $R$  repeated measurements. Applying the Law of Total Variance, we decompose the variance of  $\bar{X}_R$  as follows:

$$\begin{aligned} \mathbb{V}(\bar{X}_R) &= \mathbb{E}_i(\mathbb{V}(\bar{X}_R | (\rho_i, Q_i, \vec{G}_i))) + \mathbb{V}_i(\mathbb{E}(\bar{X}_R | (\rho_i, Q_i, \vec{G}_i))) \\ &= \frac{1}{R} \mathbb{E}_i(\mathbb{V}(X_r | (\rho_i, Q_i, \vec{G}_i))) + \mathbb{V}_i(\mathbb{E}(X_r | (\rho_i, Q_i, \vec{G}_i))). \end{aligned} \quad (5)$$

Thus, the variance of  $\bar{X}_N(R)$  is expressed as:

$$\mathbb{V}(\bar{X}_N(R)) = \frac{t(R)}{T_0} \left( \frac{Y}{R} + Z \right), \quad (6)$$

where

$$\begin{aligned} Y &= \mathbb{E}_i(\mathbb{V}(X_r | (\rho_i, Q_i, \vec{G}_i))), \\ Z &= \mathbb{V}_i(\mathbb{E}(X_r | (\rho_i, Q_i, \vec{G}_i))). \end{aligned} \quad (7)$$

Both coefficients  $Y$  and  $Z$ , related to the statistical properties of the measurement result  $X_r$ , remain constant relative to  $R$ . In the subsequent analysis, we determine the optimal reusing times  $R$  based on whether the coefficients  $Y$  and  $Z$  are known or unknown.

## 2.2 Optimizing circuit reusing with circuit and noise information

When detailed information about the noise channel and the implemented circuits is available, it becomes feasible to precisely determine the coefficients  $Y$  and  $Z$  and subsequently calculate the optimal reusing times. Assuming the cost of implementing each circuit  $R$  times follows the model  $t(R) = \alpha + \beta R$ , where  $\alpha$  and  $\beta$  are positive constants. This cost model, referred to as the “constant-cost model,” indicates that initializing a new circuit incurs an additional cost  $\alpha$ , whereas reusing an existing one costs  $\beta$  per reuse. Under this model, the optimal reusing times, denoted by  $R^*$ , can be determined as follows.

**Theorem 1.** *If the cost of implementing a circuit  $R$  times is modeled as  $t(R) = \alpha + \beta R$ , with  $\alpha, \beta$  being constant, the optimal reusing times  $R^*$  that minimizes the variance is given by*

$$R^* = \sqrt{\frac{\alpha Y}{\beta Z}}. \quad (8)$$

The above theorem can be proved by an inequality:

$$\begin{aligned} \mathbb{V}(\bar{X}_N(R)) &= \frac{\alpha + \beta R}{T_0} \left( \frac{Y}{R} + Z \right) \\ &\geq \frac{1}{T_0} \left( \sqrt{\beta Y} + \sqrt{\alpha Z} \right)^2, \end{aligned} \quad (9)$$

where the equality holds when  $R = \sqrt{\alpha Y / \beta Z}$ . From Eq. (9), it is clear that the improvement offered by the optimal strategy, compared to naively choosing  $R = 1$ , depends on the specific values of  $Y$  and  $Z$ . In the extreme case, for example when  $Y \gg Z$  and  $\alpha Z \gg \beta Y$ , the speedup factor of optimal strategy is given by:

$$\frac{\mathbb{V}(\bar{X}_N(R=1))}{\mathbb{V}(\bar{X}_N(R^*))} = \frac{(\alpha + \beta)(Y + Z)}{\beta Y + \alpha Z + 1\sqrt{\alpha\beta YZ}} \approx \frac{Y}{Z} \gg 1. \quad (10)$$

Besides, since the reusing times must be an integer, the actual optimal reusing times takes value

$$R^* = \arg \min_{R \in \{\lceil \sqrt{\frac{\alpha Y}{\beta Z}} \rceil, \lfloor \sqrt{\frac{\alpha Y}{\beta Z}} \rfloor\}} \mathbb{V}(\bar{X}_N(R)). \quad (11)$$

In future discussions, we will default on this point without ambiguity.

The conclusion of Theorem 1 can be extended to other practical cases. Here, as an example, we take the form of  $t(R)$  as

$$t(R) = C_1 \lceil R/R_c \rceil + C_2, \quad (12)$$

which is a time cost model that we observed in experiments as elaborated in Section 4. The terms  $C_1$  and  $C_2$  are both positive. In this case, the variance is given by

$$\mathbb{V}(\bar{X}_N(R)) = \frac{C_1 \lceil R/R_c \rceil + C_2}{T_0} \left( \frac{Y}{R} + Z \right). \quad (13)$$

A direct calculation indicates that in this case, the optimal reusing times takes value

$$R^* = \arg \min_{R \in \{\lceil \sqrt{\frac{C_2 Y}{C_1 Z R_c}} \rceil R_c, \lfloor \sqrt{\frac{C_2 Y}{C_1 Z R_c}} \rfloor R_c\}} \mathbb{V}(\bar{X}_N(R)). \quad (14)$$

More general cases of  $t(R)$ , including only the bound instead of concrete forms of  $t(R)$  known, are discussed in Appendix B.

### 2.3 Optimizing circuit reusing without circuit and noise information

While the above analyses show that the solution to Question 1 depends on the values of  $Y$  and  $Z$ , the two coefficients depend on both the distribution of the executed circuits and the specific noise characteristics of quantum operations. In practice, fully characterizing noise can be challenging, rendering the precise values of  $Y$  and  $Z$  often unavailable and making it difficult to pinpoint the exact optimal reusing times  $R^*$ . However, the cost model  $t(R)$ , typically determined by analyzing experimental runtime, is usually accessible. With this information at hand, it is feasible to choose a value of  $R$  that ensures the variance remains below a threshold.

Let us consider a scenario where we have prior knowledge of the cost bounds for implementing circuits multiple times:

$$\alpha_l + \beta_l R \leq t(R) \leq \alpha_u + \beta_u R, \quad \forall R \in \mathbb{Z}^+, \quad (15)$$

where  $\alpha_l$ ,  $\alpha_u$ ,  $\beta_l$ , and  $\beta_u$  are positive constants, indicating that the actual cost model  $t(R)$  falls between two such constant-cost models. Given this bounded cost model, we propose a near-optimal solution for Question 1, which does not require prior knowledge of circuit specifics or noise characteristics.

**Theorem 2.** *In a quantum learning task that averages across various circuits represented as  $\{\rho_i, Q_i, \vec{G}_i\}$ , the calculated value  $\bar{X}_N(R) = \frac{1}{NR} \sum_{i,r} X_r(\rho_i, Q_i, \vec{G}_i)$  is derived after reusing each of the  $N$  distinct circuits  $R$  times, where  $X_r$  represents the  $r$ -th one-shot measurement result. Assume that the total time cost satisfies  $T_0 = Nt(R)$  with  $\alpha_l + \beta_l R \leq t(R) \leq \alpha_u + \beta_u R$ . Within fixed total time cost, we have*

$$\mathbb{V}(\bar{X}_N(R_0)) \leq \left( \frac{\alpha_u}{\alpha_l} + \frac{\beta_u}{\beta_l} \right) \mathbb{V}(\bar{X}_N(R^*)), \quad (16)$$

where  $R_0 = \frac{\alpha_l}{\beta_l}$ , and  $R^*$  represents the optimal reusing times that minimizes  $\mathbb{V}(\bar{X}_N(R))$ .

*Proof.* Through the variance form in Eq. (6), we obtain that

$$\begin{aligned} \mathbb{V}(\bar{X}_N(R^*)) &\geq \min_R \left( \frac{\beta_l R + \alpha_l}{T_0} \left( \frac{Y}{R} + Z \right) \right) \\ &\geq \frac{\beta_l Y + \alpha_l Z}{T_0}. \\ \mathbb{V}\left(\bar{X}_N\left(\frac{\alpha_l}{\beta_l}\right)\right) &\leq \frac{\beta_u \frac{\alpha_l}{\beta_l} + \alpha_u}{T_0} \left( \frac{Y \beta_l}{\alpha_l} + Z \right) \\ &= \left( \frac{\alpha_u}{\alpha_l} + \frac{\beta_u}{\beta_l} \right) \frac{\beta_l Y + \alpha_l Z}{T_0}, \end{aligned} \quad (17)$$

By comparing the two variances, we prove our result.  $\square$

The following text refers to the near-optimal solution as  $R_0$ . When we return to the case of the constant-cost model, it is straightforward to develop a relaxed version as outlined below.

**Corollary 1.** *When the cost to implement each circuit  $R$  times is given by  $t(R) = \alpha + \beta R$  with  $\alpha$  and  $\beta$  being positive constants,  $R_0 = \frac{\alpha}{\beta}$  deems as a 2-optimal solution to Question 1, i.e.*

$$\mathbb{V}(\bar{X}_N(R_0)) \leq 2\mathbb{V}(\bar{X}_N(R^*)). \quad (18)$$

Note that the value of the optimal  $R^*$  may significantly differ from the  $\left(\frac{\alpha_u}{\alpha_l} + \frac{\beta_u}{\beta_l}\right)$ -optimal value  $R_0 = \frac{\alpha_l}{\beta_l}$  in certain scenarios. Nonetheless, the balanced strategy  $R_0$  in Theorem 2 ensures that the overall variance remains within acceptable limits, which is particularly advantageous in situations where we have no information about the form of the noise channel. More importantly, the applicability of this theorem encompasses various quantum learning and information processing methodologies employing random circuit and reusing techniques, even containing cases where the underlying sampled circuits are unknown [14, 15].

### 3 Optimal reusing times in standard RB

#### 3.1 Standard RB with circuit reusing

Here, we apply our results on optimal reusing times to RB [11], an efficient tool for assessing the quality of quantum gates, robust against state preparation and measurement (SPAM) errors. We focus on the standard RB protocol [12, 21], which evaluates the average fidelity of gates in a 2-design gate set such as the Clifford group [21, 28]. The fidelity of a gate  $G$ , denoted as  $\mathcal{G}(\rho) = G\rho G^\dagger$ , is defined by:

$$F_{avg}(\tilde{\mathcal{G}}, \mathcal{G}) := \int d\phi \operatorname{Tr}(\mathcal{G}(|\phi\rangle\langle\phi|)\tilde{\mathcal{G}}(|\phi\rangle\langle\phi|)), \quad (19)$$

where  $\tilde{\mathcal{G}}$  is the noisy implementation of  $\mathcal{G}$  and  $d\phi$  represents the uniform Haar measure over pure quantum states.

In RB protocols, the noise models of quantum gates are typically assumed to be Markovian and gate-independent [29]. Here, we adhere to this assumption, suggesting that the practical implementation of a gate  $\mathcal{G}_i$  from the gate set  $\mathbf{G}$  is represented by  $\tilde{\mathcal{G}}_i = \Lambda\mathcal{G}_i$ , where  $\Lambda$  is a completely positive trace-preserving map acting as post-gate noise. Given that  $\mathcal{G}$  is unitary,  $F_{avg}(\tilde{\mathcal{G}}, \mathcal{G})$  equals  $F_{avg}(\Lambda, \mathcal{I})$ , and we denote it as  $F_{avg}(\Lambda)$  in subsequent discussions. Additional details on the RB process and related mathematical tools are presented in Appendix A.

In short, to estimate the average fidelity of the gate set  $\mathbf{G}$  by standard RB, we sample  $N$  different gate sequences  $\vec{G}_i = \{G_1, G_2, \dots, G_m\}$  with length  $m$ , together with a global inverse gate  $G_{inv} = G_1^\dagger G_2^\dagger \dots G_m^\dagger$  for each gate sequence. Then we apply these gates to an initial state  $\rho$ , and implement positive operator-valued measure (POVM) measurement  $\{Q, \mathbb{I} - Q\}$  on the final state

$$\rho_f = \mathcal{G}_{inv}\Lambda\mathcal{G}_m \dots \mathcal{G}_2\Lambda\mathcal{G}_1(\rho). \quad (20)$$

Here, we adopt the convention of ignoring the noise channel of the inverse gate, as in other RB literature. This change does not influence the analysis of obtaining fidelity. One can view this omitted noise channel to be absorbed by measurements. With circuit reusing strategy, we repeat implementing the circuit  $(\rho, Q, \vec{G}_i)$   $R$  times with fixed  $\rho$  and  $Q$ . The  $r$ -th measurement result can be represented by a random variable

$$X_r(m, \vec{G}_i) = \begin{cases} 0, & \text{w.p. } 1 - \operatorname{Tr}(Q\rho_f) \\ 1, & \text{w.p. } \operatorname{Tr}(Q\rho_f) \end{cases}. \quad (21)$$

Averaging over  $R$  results, we obtain the survival probability associated with  $\vec{G}_i$ ,

$$\bar{X}_R(m, \vec{G}_i) = \frac{1}{R} \sum_{r=1}^R X_r(m, \vec{G}_i). \quad (22)$$



Taking the average for  $N$  different gate sequences, we obtain our final estimator,

$$\bar{X}_N(m) = \frac{1}{N} \sum_{i=1}^N \bar{X}_R(m, \vec{G}_i). \quad (23)$$

The flowchart is shown in Fig. 2.

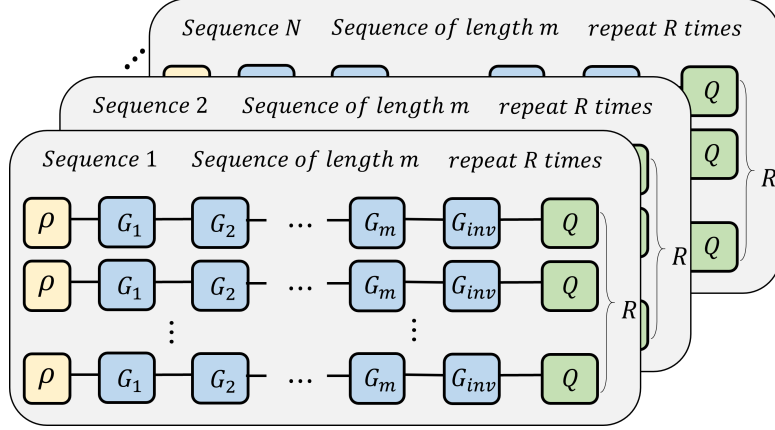


Figure 2: The flowchart illustrates the process for the standard RB protocol with quantum circuits reusing. Each layer represents a gate sequence repeating  $R$  times, and it requires  $N$  different sequences.

Utilizing the 2-design property [21, 28] of the target gate set in standard RB, the expectation value of the estimator  $\bar{X}_N(m)$  has a single exponential form,

$$\mathbb{E}(\bar{X}_N(m)) = af(\Lambda)^m + b, \quad (24)$$

where  $a$  and  $b$  are coefficients that absorb SPAM errors,  $f(\Lambda)$  represents the quality parameter associated with the target average fidelity,

$$F_{avg}(\Lambda) = f(\Lambda) + \frac{1 - f(\Lambda)}{d}, \quad (25)$$

where  $d = 2^n$  represents the system size. By assigning different sequence lengths  $m$ , one could obtain quality parameter  $f(\Lambda)$  by fitting, thus assessing the average fidelity  $F_{avg}(\Lambda)$ .

### 3.2 Theoretical analyses

Note that the variance of the fidelity estimation is positively correlated with the variance of the estimator  $\bar{X}_N(m)$  [30], hence one can optimize the variance  $\mathbb{V}(\bar{X}_N(m))$  to gain more accurate estimation for fidelity. Based on Theorem 1, to obtain the optimal reusing times  $R^*$  to minimize  $\mathbb{V}(\bar{X}_N(m))$ , we only need to evaluate the terms  $Y$  and  $Z$  in Eq. (7) according to the form of the measurement outcome  $X_r$ . In standard RB,  $X_r$  takes 1 with probability  $\text{Tr}(Q\mathcal{G}_{inv}\Lambda\mathcal{G}_m \dots \mathcal{G}_2\Lambda\mathcal{G}_1(\rho))$  and 0 otherwise, as shown in Eq. (21). Applying Theorem 1 to standard RB, we obtain the following corollary.

**Corollary 2.** *In standard RB with circuit reusing, the values of  $Y$  and  $Z$  as defined in Eq. (7) are computed by:*

$$\begin{aligned} Y &= A - B \\ Z &= B - A^2 \end{aligned} \quad (26)$$



where:

$$\begin{aligned}
A &:= \mathbb{E}_{\vec{G}} (\langle\langle Q | \mathcal{G}_{inv} \Lambda \mathcal{G}_m \dots \mathcal{G}_2 \Lambda \mathcal{G}_1 | \rho \rangle\rangle) \\
&= f(\Lambda)^m \text{Tr}(Q\rho) + \frac{1 - f(\Lambda)^m}{d} \text{Tr}(Q) \\
&= \frac{d \cdot \text{Tr}(Q\rho) - \text{Tr}(Q)}{d} \cdot \left( \frac{d \cdot F_{avg}(\Lambda) - 1}{d - 1} \right)^m + \frac{1}{d} \text{Tr}(Q), \\
B &:= \mathbb{E}_{\vec{G}} \left( \langle\langle Q | \mathcal{G}_{inv} \Lambda \mathcal{G}_m \dots \mathcal{G}_2 \Lambda \mathcal{G}_1 | \rho \rangle\rangle^2 \right).
\end{aligned} \tag{27}$$

Furthermore, when the cost of implementing a circuit  $R$  times is given by  $t(R) = \alpha + \beta R$  with positive constants  $\alpha$  and  $\beta$ , the optimal value of reusing times is determined by:

$$R^* = \sqrt{\frac{\alpha(A - B)}{\beta(B - A^2)}}. \tag{28}$$

*Proof.* By direct substitution of Eq. (21) into  $Y$  and  $Z$ , we have

$$\begin{aligned}
\mathbb{E}(X_r | \vec{G}_i) &= \langle\langle Q | \mathcal{G}_{inv} \Lambda \mathcal{G}_m \dots \mathcal{G}_2 \Lambda \mathcal{G}_1 | \rho \rangle\rangle \\
&= \text{Tr}(Q\rho_f).
\end{aligned} \tag{29}$$

$$\begin{aligned}
\mathbb{V}(X_r | \vec{G}_i) &= \mathbb{E}(X_r | \vec{G}_i) - \mathbb{E}(X_r | \vec{G}_i)^2 \\
&= \text{Tr}(Q\rho_f) - (\text{Tr}(Q\rho_f))^2.
\end{aligned} \tag{30}$$

$$\begin{aligned}
Y &= \mathbb{E}_i(\mathbb{V}(X_r | \vec{G}_i)) \\
&= \mathbb{E}_i(\text{Tr}(Q\rho_f) - (\text{Tr}(Q\rho_f))^2) \\
&= \mathbb{E}_i(\text{Tr}(Q\rho_f)) - \mathbb{E}_i((\text{Tr}(Q\rho_f))^2) \\
&= A - B.
\end{aligned} \tag{31}$$

$$\begin{aligned}
Z &= \mathbb{V}_i(\mathbb{E}(X_r | \vec{G}_i)) \\
&= \mathbb{V}_i(\text{Tr}(Q\rho_f)) \\
&= \mathbb{E}_i((\text{Tr}(Q\rho_f))^2) - (\mathbb{E}_i(\text{Tr}(Q\rho_f)))^2 \\
&= B - A^2.
\end{aligned} \tag{32}$$

Here,

$$A = \mathbb{E}_i(\text{Tr}(Q\rho_f)), \tag{33}$$

and

$$B = \mathbb{E}_i((\text{Tr}(Q\rho_f))^2). \tag{34}$$

Hence,

$$\begin{aligned}
\mathbb{V}(\bar{X}_R(m, \vec{G}_i)) &= \frac{Y}{R} + Z \\
&= \frac{A - B}{R} + B - A^2.
\end{aligned} \tag{35}$$

Considering that the circuit sequences  $\vec{G}_i$  are sampled uniformly and independently, it follows that  $\mathbb{V}(\bar{X}_N(m, \vec{G}_i)) = \frac{1}{N} \mathbb{V}(\bar{X}_R(m, \vec{G}_i))$ . When combined with the Eq. (2), we have

$$\begin{aligned}
\mathbb{V}(\bar{X}_N(m)) &= \frac{\alpha + \beta R}{T_0} \mathbb{V}(\bar{X}_R(m, \vec{G}_i)) \\
&= \frac{1}{T_0} \left( \beta(A - B) + \alpha(B - A^2) + \beta R(B - A^2) + \frac{\alpha(A - B)}{R} \right) \\
&\geq \frac{1}{T_0} \left( \sqrt{(\beta(A - B) + \alpha(B - A^2))} \right)^2,
\end{aligned} \tag{36}$$

with

$$R^* = \sqrt{\frac{\alpha(A - B)}{\beta(B - A^2)}}, \quad (37)$$

to get the minimum variance.

As for  $A$ , utilizing the 2-design property of gate set  $\mathbf{G}$ , we can rewrite  $A$  as

$$\begin{aligned} A &= \mathbb{E}_{\vec{G}} \left( \langle\langle Q | (\mathcal{G}_{inv} \Lambda \mathcal{G}_m \dots \mathcal{G}_1) \dots ((\mathcal{G}_2 \mathcal{G}_1)^\dagger \Lambda (\mathcal{G}_2 \mathcal{G}_1)) (\mathcal{G}_1^\dagger \Lambda \mathcal{G}_1) | \rho \rangle\rangle \right) \\ &= \langle\langle Q | \mathbb{E} \prod_{i=1}^m (\mathcal{G}_i^\dagger \Lambda \mathcal{G}_i) | \rho \rangle\rangle \\ &= \text{Tr}(Q \Lambda_{dep}^m(\rho)), \end{aligned} \quad (38)$$

where  $\Lambda_{dep}$  is the depolarizing channel with the same fidelity as  $\Lambda$

$$\Lambda_{dep}(\rho) = f(\Lambda)\rho + \frac{1 - f(\Lambda)}{d} \mathbb{I}, \quad (39)$$

and  $f(\Lambda)$  is the quality parameter associated with the average fidelity of  $\Lambda$  as Eq. (25). Further direct calculations complete the proof.  $\square$

To analyze the explicit formula of  $B$ , we rewrite it as

$$\begin{aligned} B &= \mathbb{E}_{\vec{G}} (\text{Tr}(Q \rho_f)^2) \\ &= \mathbb{E}_{\vec{G}} \left( \langle\langle Q^{\otimes 2} | \mathcal{G}_{inv}^{\otimes 2} \Lambda^{\otimes 2} \mathcal{G}_m^{\otimes 2} \dots \mathcal{G}_2^{\otimes 2} \Lambda^{\otimes 2} \mathcal{G}_1^{\otimes 2} | \rho^{\otimes 2} \rangle\rangle \right) \\ &= \langle\langle Q^{\otimes 2} | \mathcal{T}(\Lambda^{\otimes 2})^m | \rho^{\otimes 2} \rangle\rangle, \end{aligned} \quad (40)$$

where

$$\mathcal{T}(\Lambda^{\otimes 2}) = \mathbb{E}_{G \in \mathbf{G}} (\mathcal{G}_i^{\dagger \otimes 2} \Lambda^{\otimes 2} \mathcal{G}_i^{\otimes 2}) \quad (41)$$

is a *twirling* of quantum channel  $\Lambda^{\otimes 2}$  over the two-copy representation of  $\mathbf{G}$ . When the target gate set is the  $n$ -qubit Clifford group  $\mathbf{C}_n$ , knowing the structure of the two-copy representation of the multi-qubit Clifford group  $\mathcal{G}^{\otimes 2}$  [31] allows us to decompose  $\mathcal{T}(\Lambda^{\otimes 2})$  into several terms. One can then express  $B$  as a linear combination of exponential-decay functions with respect to  $m$ . Since there exist equivalent representations in the irreducible decomposition of the two-copy representation of the multi-qubit Clifford group,  $B$  includes matrix exponential-decay terms for general noise channel  $\Lambda$  [31]. In the next subsection, we numerically evaluate  $B$  and the optimal reusing times for different noise models.

It is worth noting that the result for  $R^*$  described in Corollary 2 can be easily generalized to the variations of RB protocols. For a given RB protocol, once detailed procedures and the benchmarking group are determined, one can always represent the  $r$ -th time measurement result under gate sequence  $\vec{G}_i$  as a random variable  $X_r(m, \vec{G}_i)$ . Then one can further obtain the value of  $Y$ ,  $Z$ , and optimal reusing times  $R^*$  by substituting  $X_r(m, \vec{G}_i)$  in Eq. (7). Interestingly, for a simple variation of standard RB, named difference RB [32], we can derive an analytical lower bound of the optimal reusing times  $R^*$ . This lower bound is a function of the fidelity and unitarity [33] of the noise channel  $\Lambda$ . For brevity, we omit the result here.

### 3.3 Numerical simulations

Let us first consider a practical experimental condition in which we possess the relaxation and decoherence times of the quantum gates [34]. These two types of times correspond to amplitude-damping and phase-damping noises, respectively. These two noise types are crucial for characterizing noise encountered in the physical realization of quantum systems. The amplitude-damping channel  $\Lambda_a$  describes an excited multi-level atomic state's decay due to the spontaneous emission of a photon, and the phase-damping channel  $\Lambda_p$ , also known as the dephasing channel, illustrates the decay of coherence over time due to external perturbations [34]. The single-qubit amplitude-damping channel is

$$\Lambda_a(\rho) = K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger, \quad K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{p} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{1-p} \\ 0 & 0 \end{pmatrix}, \quad (42)$$

and the single-qubit phase-damping channel is

$$\Lambda_p(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger, \quad E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{p} \end{pmatrix}, \quad E_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad (43)$$

where parameter  $p \in [0, 1]$  is related to the noise strength. Deriving an analytical expression for parameter  $B$  in the single-qubit case is theoretically feasible but limited to simple noise models and not easily generalizable. It is challenging to extend it to multi-qubit systems, so we used numerical simulations to estimate  $B$  and validate our results.

In our simulation, we take a composite noise channel comprising a local amplitude-damping channel  $\Lambda_{la}$  with parameter  $p_1$  alongside a local phase-damping channel  $\Lambda_{lp}$  with parameter  $p_2$ ,

$$\Lambda_{comp} = \Lambda_{lp}(p_2) \circ \Lambda_{la}(p_1), \quad (44)$$

where

$$\begin{aligned} \Lambda_{la} &= \Lambda_a^{\otimes n}, \\ \Lambda_{lp} &= \Lambda_p^{\otimes n}. \end{aligned} \quad (45)$$

Consider the case that implementing each circuit  $R$  times cost  $t(R) = R + 4$ . In this case,  $R_0 = 4$  stands as a 2-optimal value and the optimal reusing times  $R^* = \sqrt{\frac{4(A-B)}{B-A^2}}$  according to Corollary 1 and 2, respectively. Recall that once the noise channel form is determined, the value of  $A$  can be evaluated directly via the noise channel fidelity, and the value of  $B$  is a quantity averaged over multiple random Clifford gate sequences. We uniformly sample 50000 distinct Clifford circuits at random for each predefined length  $m$  to simulate the value of  $B$ . Subsequently, we compare the variance  $\mathbb{V}(\bar{X}_N)$  across different circuit reuse strategies: the optimal one  $R^*$ , the 2-optimal one  $R_0 = 4$  and the naive one  $R_{naive} = 1$ .

The value of  $R^*$  under noise channel  $\Lambda_{comp}$  across various noise parameters is illustrated in Fig. 3a and Fig. 3b. One can observe that generally, when noise level increases, i.e., noise parameter  $p_i$  becomes smaller, the value of  $R^*$  tends to reduce. This is because when the noise channel incurs a larger deviation, more randomized circuit twirling is necessary to gather adequate information concerning the noise channel, consequently leading to less circuit reusing. We also conduct a comparison between the variance of estimators for optimal  $R^*$ , 2-optimal  $R_0$ , and that for naive  $R_{naive} = 1$ , as illustrated in Fig. 3c and Fig. 3d. It can be seen that the variance for  $R_0$  does not overpass the minimum variance 2 times,

consistent with our analysis of Corollary 1. Moreover, both the optimal and 2-optimal strategies demonstrate great improvement relative to the naive selection of  $R_{naive} = 1$ . We also conduct simulations over other typical noise channels and analyze their performance, which can be found in Appendix C.

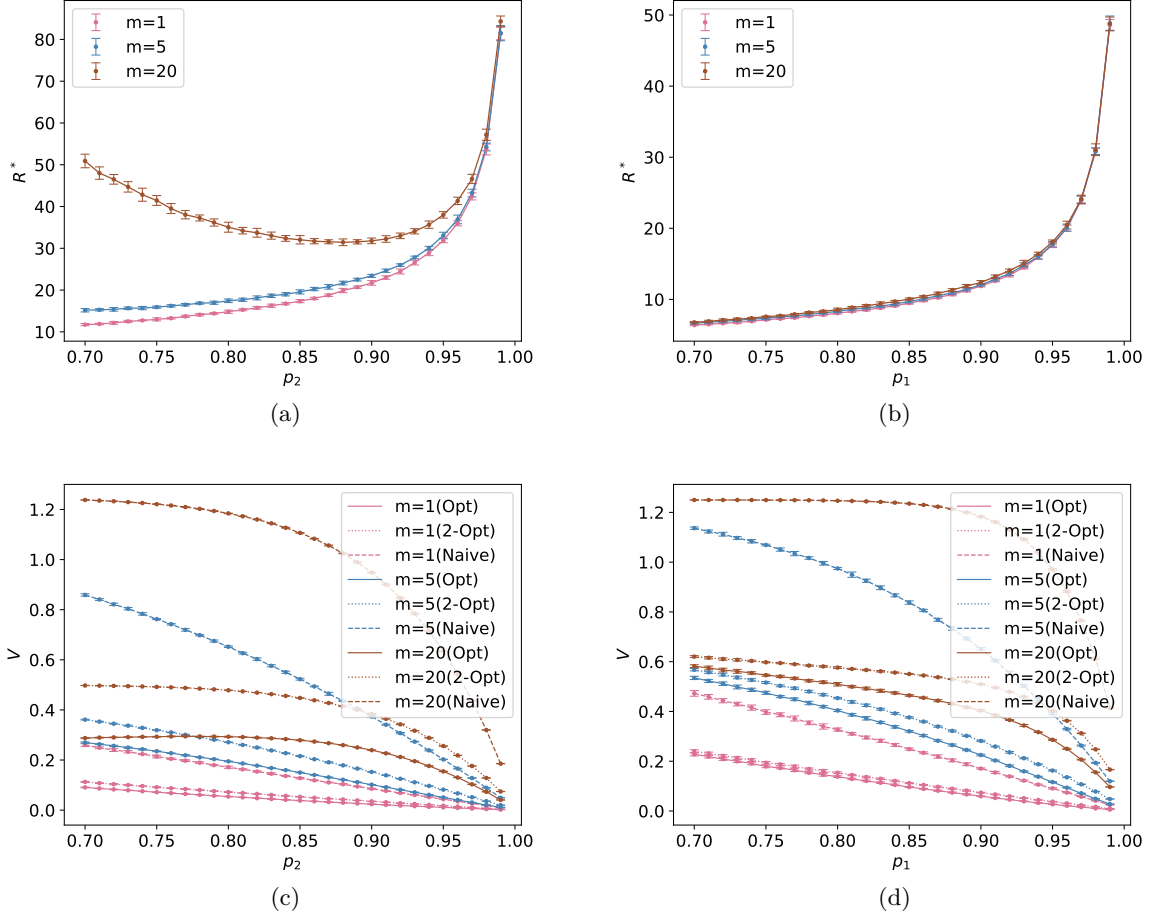


Figure 3: These figures illustrate the simulation results for the standard RB protocol under the composite channel  $\Lambda_{comp}$  defined by Eq. (44). The SPAM specifications are assigned as projective measurement  $Q = |0\rangle\langle 0|$  and initial state  $\rho = |0\rangle\langle 0|$ . Figure (a) presents the optimal reusing times  $R^*$  under the composite channel with fixed amplitude-damping parameter  $p_1 = 0.999$ , while Figure (b) shows  $R^*$  with fixed phase-damping parameter  $p_2 = 0.99$ . Figure (c) compares the variance across the naive reusing parameter  $R_{naive} = 1$ , the 2-optimal one  $R_0 = 4$  and the optimal one  $R^*$ , when the amplitude-damping parameter  $p_1 = 0.999$ , while figure (d) shows those with fixed phase-damping parameter  $p_2 = 0.99$ . We include error bars in the figures, representing twice the standard error of the mean.

## 4 Experimental results

This section presents experimental results demonstrating the application of the circuit reusing strategy in RB. We conducted a 2-qubit standard RB experiment on a superconducting quantum computing platform [27], employing various reusing times  $R$ . The set of reusing times in our experiment is presented below:

$$\mathbf{S} = \{2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000\}. \quad (46)$$

For each value of  $R$ , we implement  $N$  distinct gate sequences, detailed in Table 1.

| $R$       | 2       | 5       | 10     | 20     | 50     | 100    | 200    | 500    | 1000   | 2000   | 5000   | 10000  | 20000  |
|-----------|---------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $N$       | 300000  | 100000  | 50000  | 30000  | 10000  | 10000  | 8000   | 5000   | 3000   | 1800   | 600    | 300    | 200    |
| $T$ (s)   | 52550.3 | 17523.9 | 8774.2 | 5292.4 | 1769.6 | 1829.5 | 1787.7 | 1719.0 | 1643.1 | 1720.3 | 1311.5 | 1269.0 | 1663.1 |
| $T_0$ (s) | 53250.0 | 17750.0 | 8875.0 | 5325.0 | 1775.0 | 1775.0 | 1748.0 | 1707.5 | 1639.5 | 1721.7 | 1311.9 | 1271.0 | 1667.3 |
| $T_0/T$   | 1.013   | 1.013   | 1.011  | 1.006  | 1.003  | 0.970  | 0.978  | 0.993  | 0.998  | 1.001  | 1.000  | 1.002  | 1.003  |

Table 1: Comparison between the practical experimental runtime  $T(N, R)$  and the estimated runtime  $T_0(N, R)$  by our cost model  $T_0(N, R) = C_1 N \lceil R/R_c \rceil + C_2 N$  with parameters  $(C_1, C_2, R_c) = (0.0410s, 0.1365s, 100)$ .

The experimental runtime data, summarized in Table 1, inspires us to adopt a ladder-like model for the total time expense  $T_0$ , dependent on  $R$ :

$$T_0(N, R) = C_1 N \lceil R/R_c \rceil + C_2 N, \quad (47)$$

where the coefficients are determined to be  $(C_1, C_2, R_c) = (0.0410s, 0.1365s, 100)$ . Accordingly, the cost to implement each circuit  $R$  times is given by:

$$t(R) = C_1 \lceil R/R_c \rceil + C_2, \quad (48)$$

a model previously employed in the optimization analysis in Section 2.2. To verify the accuracy of this cost model, we compare the experimental runtime per circuit  $T/N$  against the values calculated using our model  $T_0/N$ , as shown in Fig. 4a.

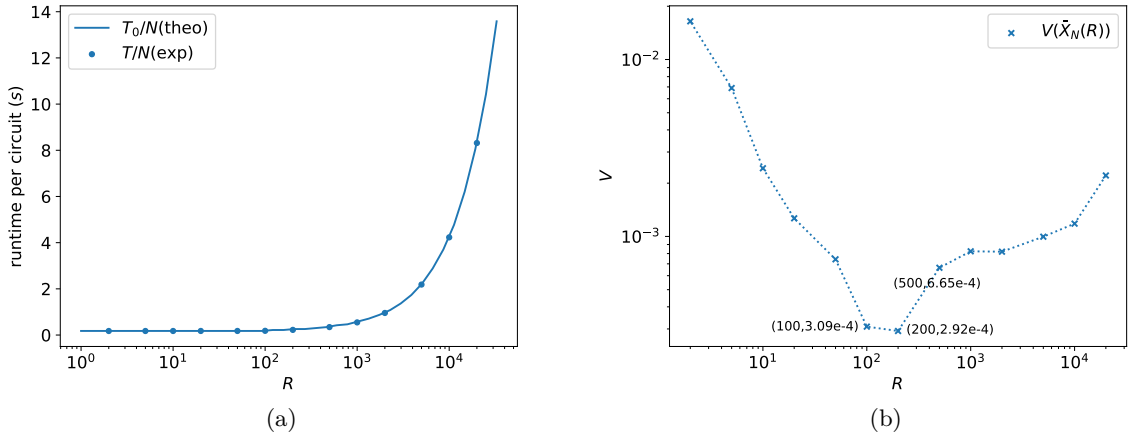


Figure 4: (a) This plot compares the experimental runtime per circuit,  $T/N$ , for different reusing times  $R$ , with the predicted costs from our model,  $T_0/N$ . The alignment of these values validates our cost model. (b) This plot displays the variance of the average measurement results,  $\bar{X}_N(R)$ , across various reusing times  $R$  and different numbers of unique gate sequences  $N'$  as specified in Table 2. Each sequence is executed over a circuit length of  $m = 30$ . The reusing times with the experimental minimum variance is 200, and the corresponding variance is  $2.92 \times 10^{-4}$ . We also label the variances associated with the reusing times 100 and 500 in the figure.

From an experimental standpoint, the physical intuition behind this cost model presented in Eq. (47) is clear. The term  $C_2 N$  represents the computational time expense required for generating each circuit, including the computation and decomposition of the inverse gate as well as the random selection of each gate sequence. The term  $C_1 N \lceil R/R_c \rceil$

represents the hardware execution time, which includes circuit implementation, data packing and transmission, and other associated tasks. Among these, the dominant time cost arises from data packing and transmission. In our quantum computing platform, data packing and transmission operations are performed in batches, with each batch containing  $R_c$  units. After  $R_c$  iterations of measurement, the results are transmitted to the classical computer as a single batch. This batching strategy is commonly used in classical computer architecture to amortize the computational overhead of specific operations across multiple instances, thereby improving system efficiency. Consequently, the factor  $\lceil R/R_c \rceil$  indicates the number of “batches” into which the data transfer between quantum devices and classical computers is divided, while  $C_1$  represents the cost per batch. In contrast, the time required for circuit implementation is relatively short and can be omitted.

For each reusing times  $R \in \mathbf{S}$ , we compute the variance of the average survival probability as depicted in Fig. 4b. To ensure uniform experimental costs across different values of  $R$ , we employ  $N'$  distinct gate sequences for each value of  $R$  to calculate the survival probability, where the values of  $N'$  are detailed in Table 2. This process is repeated 10 times to obtain 10 distinct survival probabilities for each  $(R, N')$ . Fig. 4b illustrates the variance among these 10 values. Specifically, the variances of the survival probabilities are  $3.09 \times 10^{-4}$ ,  $2.92 \times 10^{-4}$ , and  $6.65 \times 10^{-4}$  for  $R = 100$ , 200, and 500, respectively. These values represent the three smallest variances observed in the experiments, suggesting that the experimental optimal reusing times  $R_{exp}^*$  is approximately 200, and lies within the range  $[100, 500]$ .

| $R$   | 2      | 5      | 10     | 20     | 50      | 100     | 200     | 500     | 1000    | 2000    | 5000   | 10000  | 20000  |
|-------|--------|--------|--------|--------|---------|---------|---------|---------|---------|---------|--------|--------|--------|
| $N'$  | 715    | 715    | 715    | 715    | 715     | 715     | 581     | 372     | 232     | 133     | 58     | 30     | 15     |
| $V$   | 0.0164 | 0.0069 | 0.0024 | 0.0013 | 0.00074 | 0.00031 | 0.00029 | 0.00067 | 0.00082 | 0.00082 | 0.0010 | 0.0012 | 0.0022 |
| $T_0$ | 125.05 | 127.09 | 126.82 | 127.21 | 126.79  | 127.04  | 126.95  | 126.91  | 126.91  | 126.91  | 126.91 | 126.91 | 126.91 |

Table 2: Experimental specifications for different reusing times  $R$  with corresponding distinct gate sequences  $N'$  to ensure uniform experimental runtime. Each set of measurements is repeated 10 times to evaluate the variance in survival probabilities shown in Fig. 4b.

Utilizing the condition that  $R/R_c \leq \lceil R/R_c \rceil \leq (R + R_c - 1)/(R_c)$ , we derive that

$$\frac{C_1}{R_c}R + C_2 \leq t(R) \leq \frac{C_1}{R_c}R + C_2 + C_1 \left(1 - \frac{1}{R_c}\right). \quad (49)$$

Based on Theorem 2, we determine that  $R_0 = \frac{C_2 R_c}{C_1}$  is a  $\left(2 + \frac{C_1}{C_2} \left(1 - \frac{1}{R_c}\right)\right)$ -optimal value. Using the experimental coefficients  $C_1, C_2$ , and  $R_c$ , we find  $R_0 = 333$  to be a 2.3-optimal value. This theoretical finding is consistent with experimental data from  $R = 200$  and  $R = 500$ , as illustrated in Fig. 4b.

Furthermore, we utilize the mean of the experimental measurement outcomes to approximate the expectations  $A$  and  $B$  as defined in Corollary 2, determining the theoretical

optimal reusing times  $R^*$ .

$$\begin{aligned}
A &= \mathbb{E}_{\vec{G}} (\langle\langle Q | \mathcal{G}_{inv} \Lambda \mathcal{G}_m \dots \mathcal{G}_2 \Lambda \mathcal{G}_1 | \rho \rangle\rangle) \\
&= \mathbb{E}_{\vec{G}} \Pr(0 | \vec{G}) \\
&\approx \overline{\Pr(0)} \\
&= 0.1482, \\
B &= \mathbb{E}_{\vec{G}} (\langle\langle Q | \mathcal{G}_{inv} \Lambda \mathcal{G}_m \dots \mathcal{G}_2 \Lambda \mathcal{G}_1 | \rho \rangle\rangle^2) \\
&= \mathbb{E}_{\vec{G}} \Pr(0 | \vec{G})^2 \\
&\approx \overline{\Pr(0)^2} \\
&= 0.0248, \\
Y &= A - B = 0.1234, \\
Z &= B - A^2 = 0.0028, \\
\sqrt{\frac{C_2 Y}{C_1 Z R_c}} &= 1.20, \\
R^* &\in \{100, 200\}.
\end{aligned} \tag{50}$$

Here,  $\Pr(0 | \vec{G})$  denotes the probability of obtaining the measurement result 0 when implementing the gate sequence  $\vec{G}$ ;  $\overline{\Pr(0)}$  represents the average probability of obtaining 0 across all experimentally realized gate sequences; and  $\overline{\Pr(0)^2}$  is the average of the squared probabilities of obtaining 0 for all experimentally realized gate sequences. This result also aligns with the experimental optimal reusing times  $R_{exp}^*$ .

Note that despite the practical experiment for RB not strictly conforming to the gate-independent and Markovian noise assumption, the interval of our theoretical optimal reusing times remains in alignment with the experimental results and provides useful insights into reducing the experimental overhead.

## 5 Conclusion and outlook

Reusing quantum circuits often presents a more cost-effective approach compared to preparing and executing new circuits. In this work, we have developed a general analytical framework to optimize the circuit reusing times  $R$  in quantum learning tasks. By employing the Law of Total Variance, we decompose the variance of outcomes into two components, enabling us to derive a near-optimal solution for determining  $R$ . This solution is remarkably versatile and applicable to various quantum learning tasks irrespective of specific noise models. Furthermore, when the noise characteristics of circuits are known, we can pinpoint the exact optimal reusing times  $R^*$ .

We experimentally implemented the standard RB protocol on a superconducting quantum computing platform to validate our theoretical framework. Utilizing the experimental runtime, we constructed a cost model specific to this platform and determined the optimal and near-optimal reusing times. The experimental data demonstrate that the initialization of a new circuit incurs a higher cost, underscoring the imperative of the circuit reusing strategy in experiments. Nonetheless, the relationship between  $R$  and the total cost is observed to be non-linear, contradicting previous assumptions in the literature. Finally, the experimentally observed optimal reusing times, which minimizes experimental variance, aligns with our theoretical predictions.



For future exploration, our variance-cost analytical framework holds the potential to optimize protocols for characterizing various quantum properties, such as entanglement [26] and coherence [35]. Our analysis thus far has assumed independent and identically distributed conditions for measurement outcomes. Future investigations could explore how our findings extend to scenarios where executed circuits exhibit correlations. Additionally, it would be valuable to extend our analysis to a wider range of RB scenarios and investigate the reusing strategy within generalized RB frameworks [29, 36]. Lastly, our results suggest revisiting similar studies in other quantum learning tasks, like thrifty shadow estimation [19, 20], where the linear model might not be applicable.

## Acknowledgements

We express our gratitude to Zhenyu Du, Chu Zhao, Zitai Xu, and Ziyi Xie for their insightful discussions. The experiments in this work are performed on the superconducting platform [27] from the University of Science and Technology of China. We especially thank Daojin Fan and Ming Gong for the experimental implementation. This work was supported by the National Natural Science Foundation of China Grant No. 12174216 and the Innovation Program for Quantum Science and Technology Grant No. 2021ZD0300804 and No. 2021ZD0300702.

## A Preliminary

### A.1 Representation theory

In this part, we recall some useful facts regarding the representations of finite groups. Consider a finite group,  $\mathbf{G}$ , and a finite-dimensional complex vector space  $V$ . Denote  $GL(V)$  as the group of invertible linear transformations of  $V$ , we can define a representation  $\phi$  of the group  $\mathbf{G}$  on the space  $V$  as a map

$$\phi : \mathbf{G} \rightarrow GL(V) : g \mapsto \phi(g), \quad (51)$$

such that

$$\phi(g)\phi(h) = \phi(gh), \quad \forall g, h \in \mathbf{G}. \quad (52)$$

Given representation  $\phi$  on  $V$ , a linear subspace  $W \subseteq V$  is called *invariant* if

$$\phi(g)w \in W, \quad \forall w \in W, \forall g \in \mathbf{G}. \quad (53)$$

The restriction of  $\phi$  to the invariant subspace  $W$  is known as a *subrepresentation* of  $\mathbf{G}$  on  $W$ . If there is a non-trivial subspace  $W$  of  $V$ , i.e.  $W \neq \{0\}, V$ , such that

$$\phi(g)W \subset W, \quad \forall g \in \mathbf{G}, \quad (54)$$

then the representation  $\phi$  is called *reducible*. If there are no non-trivial subspaces  $W$  satisfying Eq. (54), the representation  $\phi$  is called *irreducible*. Two representations  $\phi, \phi'$  of a group  $\mathbf{G}$  on spaces  $V, V'$  are called *equivalent* if there exists an invertible linear map  $T : V \rightarrow V'$  such that

$$T \circ \phi(g) = \phi'(g) \circ T, \quad \forall g \in \mathbf{G}. \quad (55)$$

Maschke's lemma states that every representation  $\phi$  of a group can be uniquely written as a direct sum of irreducible representations

$$\phi(g) \simeq \bigoplus_{\lambda \in R_G} \phi_\lambda(g)^{\oplus m_\lambda}, \quad \forall g \in \mathbf{G}, \quad (56)$$

where  $R_G$  is the index set of the irreducible representations, and  $m_\lambda$  represents the multiplicity of the equivalent irreducible representation of  $\phi_\lambda$  in  $\phi$ .

One important object in representation theory is *character*, the *character*  $\chi_\phi$  corresponding to a representation  $\phi$  of a group  $\mathbf{G}$  is defined as

$$\chi_\phi(g) = \text{Tr}(\phi(g)). \quad (57)$$

With the character, we introduce the *projector* onto irreducible representation. Let  $\phi_\lambda$  be an irreducible representation contained in  $\phi$  with character  $\chi_\lambda$ , then the projector onto the support space of  $\phi_\lambda$  can be described by

$$\mathcal{P}_{\phi_\lambda} = \frac{|\phi_\lambda|}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} \chi_\lambda(g) \phi(g), \quad (58)$$

where  $|\phi_\lambda|$  denotes the dimension of  $\phi_\lambda$ .

Define the twirl  $\mathcal{T}_\phi(M)$  of a linear map  $M : V \rightarrow V$  with respect to the representation  $\phi$  to be

$$\mathcal{T}_\phi(M) := \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} \phi(g) M \phi(g)^\dagger. \quad (59)$$

Now we can further analyze the twirl map over specific types of representations by corollary of Schur's lemma:

**Lemma 1.** *Let  $\mathbf{G}$  be a finite group and let  $\phi$  be a representation of  $\mathbf{G}$  on a complex vector space  $V$  with decomposition*

$$\phi(g) \simeq \bigoplus_{\lambda} \phi_\lambda(g), \quad \forall g \in \mathbf{G} \quad (60)$$

*into inequivalent irreducible subrepresentations  $\phi_\lambda$ , i.e.  $m_\lambda = 1$ . Then for any linear operation  $M : V \rightarrow V$ , the twirl map of  $M$  over  $\mathbf{G}$  is given by*

$$\mathcal{T}_\phi(M) = \sum_{\lambda} \frac{\text{Tr}(M \mathcal{P}_\lambda)}{\text{Tr}(\mathcal{P}_\lambda)} \mathcal{P}_\lambda, \quad (61)$$

where  $\mathcal{P}_\lambda$  is the projector onto the subspace of  $\phi_\lambda$ .

Denote  $f_\lambda = \frac{\text{Tr}(M \mathcal{P}_\lambda)}{\text{Tr}(\mathcal{P}_\lambda)}$ , we can find that,

$$\mathcal{T}_\phi(M)^m = \sum_{\lambda} f_\lambda^m \mathcal{P}_\lambda, \quad (62)$$

on account of the idempotent and mutually orthogonal properties of projectors  $\mathcal{P}_\lambda$ . Hence, to benchmark the gate set of a group  $\mathbf{G}$ , one can sample the circuit across the group with a given length  $m$ , then measure the averaged survival probabilities

$$\mathbb{E}(X(m)) = \sum_{\lambda} C_\lambda f_\lambda^m \quad (63)$$

with coefficient  $C_\lambda$  associated with SPAM terms. By applying post-processing to fit an exponential decay model, the average fidelity of the gate set can be estimated by the decay parameters  $f_\lambda$ , which is the key technique of RB [29]. However, this functional form holds only when representation  $\phi$  has no irreducible subrepresentation occurs more than once,

that is, in the  $m_\lambda = 1$  case. In the general case with multiplicity  $m_\lambda \neq 1$ , the averaged survival probability takes the form

$$X(m) \approx \sum_{\lambda} \text{Tr}(C_\lambda M_\lambda^m), \quad (64)$$

where  $M_\lambda$  is an  $m_\lambda \times m_\lambda$  real matrix that only depends on the implementation of  $\phi$  and  $C_\lambda$  is an  $m_\lambda \times m_\lambda$  matrix also associated with SPAM behavior [29]. It is important to note that the matrices  $M_\lambda$  may not be diagonalizable, rendering it impossible to determine the decay parameter associated with average fidelity by fitting in this scenario. Consequently, the majority of RB protocols concentrate primarily on scenarios where  $m_\lambda = 1$ .

## A.2 Liouville representation of quantum channel

In this part, we will introduce the Liouville representation, which is extensively utilized in discussions about quantum channels. Denote the Hilbert space for  $n$  qubits as  $\mathcal{H}$  and the set of linear operators on  $\mathcal{H}$  as  $\mathcal{L}(\mathcal{H})$ . The Liouville representation is defined on a set of the trace-orthonormal basis on  $\mathcal{L}(\mathcal{H})$ , and the most widely used basis is the normalized Pauli group

$$\mathbf{P}'_n := \left\{ \sigma_i = \frac{P_i}{\sqrt{d}} \mid P_i \in \{\mathbb{I}, X, Y, Z\}^{\otimes n} \right\}, \quad (65)$$

where  $d = 2^n$  is the system dimension. A good property of this basis is that each normalized Pauli operator is complete and satisfies

$$\text{Tr}(\sigma_i^\dagger \sigma_j) = \delta_{ij}, \quad \forall \sigma_i, \sigma_j \in \mathbf{P}'_n \quad (66)$$

under the Hilbert-Schmidt inner product. Any  $n$ -qubit operator  $O \in \mathcal{L}(\mathcal{H})$  can be decomposed over the  $4^n$  normalized Pauli operators,

$$O = \sum_{\sigma_i \in \mathbf{P}'_n} \text{Tr}(\sigma_i^\dagger O) \sigma_i. \quad (67)$$

Thus, one can reformulate  $O$  in a vector form,

$$|O\rangle\rangle = \sum_{\sigma_i \in \mathbf{P}'_n} \text{Tr}(\sigma_i^\dagger O) |\sigma_i\rangle\rangle, \quad (68)$$

where  $|\sigma_i\rangle\rangle$  in Liouville representation is a  $4^n$ -length vector with only one non-zero element 1.

Recall that quantum channels are known as completely positive and trace-preserving linear maps on  $\mathcal{L}(\mathcal{H})$ . Through the Liouville representation, any quantum channel can be represented as a matrix. The action of the channel  $\Lambda$  on an operator  $O$  is given by

$$|\Lambda(O)\rangle\rangle = \Lambda|O\rangle\rangle, \quad (69)$$

where  $\Lambda_{ij} = \langle\langle \sigma_i | \Lambda | \sigma_j \rangle\rangle = \text{Tr}(\sigma_i \Lambda(\sigma_j))$ . In the main text, what we mainly use is the action of a unitary operation  $G$  on a state  $\rho$ :

$$|G\rho G^\dagger\rangle\rangle = \mathcal{G}|\rho\rangle\rangle. \quad (70)$$

The composition of two channels can be depicted as the product of two matrices,

$$|\Lambda_2 \circ \Lambda_1(O)\rangle\rangle = \Lambda_2 \Lambda_1 |O\rangle\rangle. \quad (71)$$

Furthermore, the measurement operator can also be vectorized with the Liouville notation according to the definition of the Hilbert-Schmidt inner product. For example, the measurement probability of a state  $\rho$  on a positive operator-valued measure (POVM)  $\{F_i\}$  is described as

$$p_i = \langle\langle F_i | \rho \rangle\rangle = \text{Tr}(F_i^\dagger \rho). \quad (72)$$

### A.3 Randomized benchmarking

The following box presents the detailed procedure of standard RB with  $R$  times circuit reusing.

#### Standard RB

1. Sample a random circuit  $\vec{G}_i = \{G_1, \dots, G_m\}$  of length  $m$ , where each gate  $G_i$  is independently and uniformly sampled from the gate set  $\mathbf{G}$ .
2. Compute the global inverse gate  $G_{inv} = (G_m \cdots G_1)^{-1}$ , and add it to the end of  $\vec{G}_i$  to ensure that the entire circuit is equivalent to the identity channel  $\mathcal{I}$  in the noiseless case.
3. Apply the circuit with inverse gate to the initial state  $\rho$ , and then take POVM measurement  $\{Q, \mathbb{I} - Q\}$ .
4. Repeat step 3 for  $R$  times, the result of  $r$ -th measurement can be described as a random variable  $X_r(m, \vec{G}_i)$ , and estimate the survival probability of sequence  $\vec{G}_i$ :

$$\bar{X}_R(m, \vec{G}_i) = \frac{1}{R} \sum_r X_r(m, \vec{G}_i). \quad (73)$$

5. Repeat steps 1-4 for  $N$  distinct circuits  $\vec{G}_i$ , and estimate the overall survival probability by aggregating the result derived from each set of measurements:

$$\bar{X}_N(m) = \frac{1}{N} \sum_i \bar{X}_R(m, \vec{G}_i). \quad (74)$$

6. Repeat steps 1-5 across a sufficient range of different lengths  $m$  and fit  $\bar{X}_N(m)$  to the following decay function to get  $f(\Lambda)$ :

$$\bar{X}_N(m) = af(\Lambda)^m + b. \quad (75)$$

In summary, by random sampling, we obtain the survival probabilities of circuits for different lengths  $m$  and fit them to an exponential decay function to get the average fidelity of the gate set  $\mathbf{G}$

$$F_{avg}(\Lambda) = f(\Lambda) + \frac{1 - f(\Lambda)}{d}. \quad (76)$$

Note that the initial state  $\rho$  and POVM measurement  $\{Q, \mathbb{I} - Q\}$  are fixed.

## B Optimizing circuit reusing times with circuit and noise information for generic cost models

In this part, we discuss the optimal reusing times for generic cost models. Recall that for a quantum learning task, the variance of the estimator is described as

$$\mathbb{V}(\bar{X}_N(R)) = \frac{t(R)}{T_0} \left( \frac{Y}{R} + Z \right), \quad (77)$$

where  $Y$  and  $Z$  are coefficients associated with measurement results. Normally, the optimal reusing times can be determined by minimizing Eq. (77) once the specific form of cost model

$t(R)$  is given. Here, we discuss a more general case where we only possess prior knowledge about the bound of the cost  $t(R)$ :

$$\beta_l R + \alpha_l \leq t(R) \leq \beta_u R + \alpha_u, \quad \forall R \in \mathbb{Z}^+ \quad (78)$$

with constants  $\alpha_l, \alpha_u, \beta_l, \beta_u$  all positive. This is the case that the real cost model  $t(R)$  is bounded by two constant-cost models, as mentioned in the main text. Given the values of  $Y$  and  $Z$ , we can identify the range of the optimal solution to minimize the variance, as shown in the following theorem.

**Theorem 3.** *Given the bound of the cost model  $t(R)$ ,  $\beta_l R + \alpha_l \leq t(R) \leq \beta_u R + \alpha_u$  with  $\alpha_l, \alpha_u, \beta_l, \beta_u$  all positive, within fixed total time cost, the optimal reusing times  $R^*$  to minimize the variance  $\mathbb{V}(\bar{X}_N(R))$  satisfies*

$$R^* \in \left[ \frac{b - \sqrt{b^2 - 4ac}}{2a}, \frac{b + \sqrt{b^2 - 4ac}}{2a} \right], \quad (79)$$

with

$$\begin{aligned} a &= \beta_l Z; \\ b &= (\sqrt{\alpha_u Z} + \sqrt{\beta_u Y})^2 - (\alpha_l Z + \beta_l Y); \\ c &= \alpha_l Y. \end{aligned} \quad (80)$$

*Proof.* Since  $t(R) \leq \beta_u R + \alpha_u$ , we have

$$\mathbb{V}(\bar{X}_N(R)) \leq \frac{\beta_u R + \alpha_u}{T_0} \left( \frac{Y}{R} + Z \right). \quad (81)$$

Thus,

$$\begin{aligned} \mathbb{V}(\bar{X}_N(R^*)) &= \min_R \mathbb{V}(\bar{X}_N(R)) \\ &\leq \min_R \frac{\beta_u R + \alpha_u}{T_0} \left( \frac{Y}{R} + Z \right) \\ &= \frac{(\sqrt{\alpha_u Z} + \sqrt{\beta_u Y})^2}{T_0}. \end{aligned} \quad (82)$$

On the other hand, since  $t(R) \geq \beta_l R + \alpha_l$ , we have

$$\mathbb{V}(\bar{X}_N(R)) \geq \frac{\beta_l R + \alpha_l}{T_0} \left( \frac{Y}{R} + Z \right). \quad (83)$$

By substituting  $R^*$  into the above inequality, we get

$$\frac{\beta_l R^* + \alpha_l}{T_0} \left( \frac{Y}{R^*} + Z \right) \leq \frac{(\sqrt{\alpha_u Z} + \sqrt{\beta_u Y})^2}{T_0}. \quad (84)$$

The above inequality is equivalent to  $R^* \in \left[ \frac{b - \sqrt{b^2 - 4ac}}{2c}, \frac{b + \sqrt{b^2 - 4ac}}{2c} \right]$ , which finishes the proof.  $\square$

Theorem 3 provides a bound for the optimal reusing times once the cost model is bounded. Note that when the cost model satisfies  $t(R) = R + \alpha$  with a constant  $\alpha$ , the result of Theorem 3 becomes  $R^* = \sqrt{\frac{\alpha Y}{Z}}$ , which is consistent with Theorem 1. In this sense, Theorem 3 can be viewed as a generalization of Theorem 1.

## C Analysis of optimal reusing times under various noise channels

In experiments, we can utilize prior knowledge about the noise channel to estimate optimal reusing values. Here, we analyze the optimal reusing times  $R^*$  for standard RB by numerical simulation across typical noise channels, specifically:

$$\begin{aligned}
&\text{global depolarizing channel: } \Lambda_{dep}(\rho) = p\rho + (1-p)\frac{\mathbb{I}}{2^n}, \\
&\text{local amplitude-damping channel: } \Lambda_{la} = \Lambda_a^{\otimes n}, \\
&\text{where } \Lambda_a(\rho) = K_0\rho K_0^\dagger + K_1\rho K_1^\dagger, K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{p} \end{pmatrix}, K_1 = \begin{pmatrix} 0 & \sqrt{1-p} \\ 0 & 0 \end{pmatrix}, \\
&\text{local phase-damping noise channel: } \Lambda_{lp} = \Lambda_p^{\otimes n}, \\
&\text{where } \Lambda_p(\rho) = E_0\rho E_0^\dagger + E_1\rho E_1^\dagger, E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{p} \end{pmatrix}, E_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \\
&\text{local unitary noise channel: } \Lambda_{lz} = \Lambda_z^{\otimes n}, \text{ where } \Lambda_z(\rho) = \exp(i\theta\mathcal{Z})(\rho) = i\sin\theta\mathcal{Z}(\rho) + \cos\theta \cdot \rho,
\end{aligned} \tag{85}$$

where  $\mathcal{Z}(\rho) = Z\rho Z^\dagger$ ,  $n$  denotes the system qubit number, and  $p \in [0, 1], \theta \in [0, \pi]$  are related parameters of the channels. To cover multi-qubit noise channels, we also consider a “correlation channel”:

$$\Lambda_c = \bigotimes_{i < j} \exp(i\beta_{ij}\text{SWAP}_{ij}), \tag{86}$$

where  $\beta_{ij} \in [0, \pi]$  is the correlation parameter representing the interaction strength between qubits  $i$  and  $j$ . These models can effectively approximate the vast majority of noises encountered in practical experiments.

To depict the trend of  $R^*$  under different kinds of noise channel parameters, we consider the case that reusing an existing circuit  $R$  times cost  $t(R) = R + 4$ . In this case,  $R_0 = 4$  stands as a 2-optimal value according to Corollary 1, and the optimal reusing times reads  $R^* = \sqrt{\frac{A-B}{B-A^2}}$ . If the actual cost factor is not equal to 4 but another  $\alpha^*$ , the optimal reusing times obtained from simulations can be adjusted by multiplying the factor  $\sqrt{\alpha^*}$ , as the theoretical optimal  $R^*$  is proportional to  $\sqrt{\alpha}$  for the constant-cost model  $t(R) = R + \alpha$ . To determine  $R^*$ , one has to get the values of  $A$  and  $B$ . The value of  $A$  can be evaluated directly via noise channel fidelity. The value of  $B$  is a quantity averaged over multiple random Clifford gate sequences. We uniformly sample 50000 distinct Clifford circuits at random for each predefined length  $m$  to simulate the value of  $B$ . Then, we can observe how much the optimal  $R^*$  reusing strategy and the 2-optimal  $R_0$  reusing strategy improve compared to naively choosing  $R_{naive} = 1$ . Within this section, for the  $n$ -qubit simulations the SPAM specifications are assigned as  $\rho = |0\rangle\langle 0|^{\otimes n}$  and  $Q = |0\rangle\langle 0|^{\otimes n}$ .

### C.1 Global depolarizing noise channel

Depolarizing noise is an important environmental noise characterization for a quantum channel [34]. Under the global depolarizing noise channel, abbreviated as “depolarizing channel” without ambiguity, we observe that the optimal reusing times  $R^* \rightarrow \infty$  for the standard RB protocol. This occurs because the final state under application of arbitrary

circuit  $\vec{G} = \{G_1, G_2, \dots, G_m, G_{inv}\}$ ,

$$\begin{aligned}
\rho_f(\vec{G}) &= \mathcal{G}_{inv} \Lambda_{dep} \mathcal{G}_m \Lambda_{dep} \dots \Lambda_{dep} \mathcal{G}_1(\rho) \\
&= p \mathcal{G}_{inv} \Lambda_{dep} \mathcal{G}_m \Lambda_{dep} \dots \Lambda_{dep} \mathcal{G}_1(\rho) + (1-p) \mathcal{G}_{inv} \Lambda_{dep} \mathcal{G}_m \Lambda_{dep} \dots \Lambda_{dep} \mathcal{G}_2(\frac{\mathbb{I}}{2^n}) \\
&= p^m \mathcal{G}_{inv} \mathcal{G}_m \dots \mathcal{G}_1(\rho) + (1-p^m) (\frac{\mathbb{I}}{2^n}) \\
&= p^m(\rho) + (1-p^m) (\frac{\mathbb{I}}{2^n}),
\end{aligned} \tag{87}$$

is solely dependent on the noise channel parameter  $p$  and circuit length  $m$ , where  $\mathcal{G}_i$  represents the Liouville representation for  $G_i$ . Consequently, the quality parameter can be extracted without the need for twirling over the entire gate set. Instead, one can simply reuse a single circuit  $N$  times to estimate the noise channel and minimize the experimental cost. The underlying reason for the result  $R^* \rightarrow \infty$  is the commutative property of the global depolarizing channel with respect to quantum unitary operations. Hence, this result is not only applicable to the standard RB protocol but also to other RB protocols where the overall circuit becomes identity channel  $\mathcal{I}$  for the noiseless case [29].

However, experimental noise is typically not limited to depolarizing noise alone. To analyze the impact of depolarizing noise on the optimal reusing times  $R^*$ , we consider a composite noise channel comprising a local amplitude-damping channel with parameter  $p_1$  alongside a depolarizing channel with parameter  $p_2$ :

$$\Lambda_{comp2}(\rho) = \Lambda_{dep}(p_2) \circ \Lambda_{la}(p_1)(\rho), \tag{88}$$

with fixed amplitude-damping parameter  $p_1 = 0.99$ . The simulation results for  $R^*$  and comparison of variance across the optimal  $R^*$ , 2-optimal  $R_0$ , and naive  $R_{naive}$  are illustrated in Fig. 5. As  $p_2$  approaches 0,  $R^*$  increases since the depolarizing part becomes more dominant, and it is known to us that  $R^* \rightarrow \infty$  for the pure depolarizing channel. On the other hand, when  $p_2$  is close to 1,  $R^*$  approaches the result under pure local amplitude-damping channel with given parameter  $p_1$ .

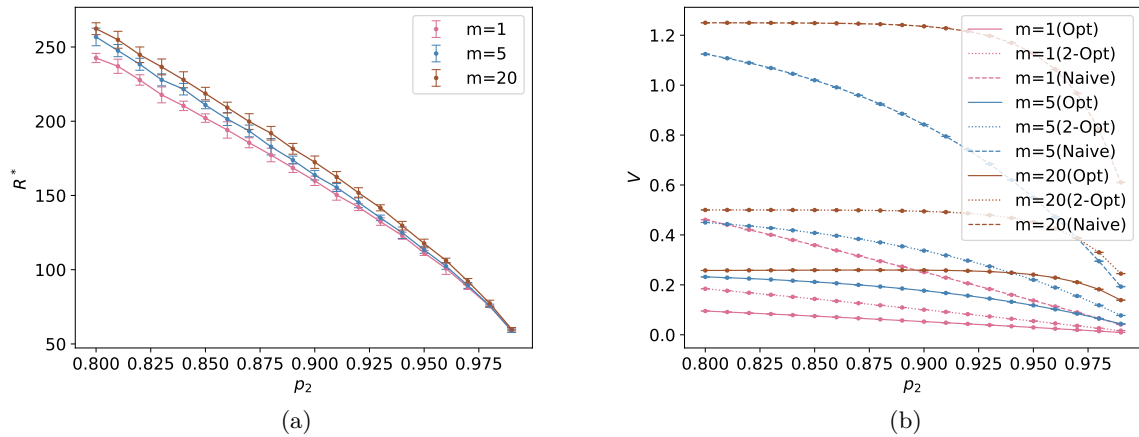


Figure 5: Figure (a) presents the optimal reusing times  $R^*$  for standard RB over composite noise channel  $\Lambda_{comp2}$  defined by Eq. (88) with parameter  $p_1 = 0.99$  and  $p_2$  varying as the x-axis, while figure (b) shows a comparison between the variance of the estimator under the optimal  $R^*$ , under the 2-optimal  $R_0 = 4$ , and under the naive  $R_{naive} = 1$  within same noise scenario.

## C.2 Local unitary noise channel

For the local unitary noise channel in standard RB, let us first consider an extreme case with the noise channel parameter  $\theta = \frac{\pi}{2}$ :  $\Lambda_{lz}(\theta = \frac{\pi}{2}) = (i\mathcal{Z})^{\otimes n}$ . The final state under this channel can be described as

$$\begin{aligned}\rho_f(\vec{G}, m) &= \mathcal{G}_{inv} \Lambda_{lz} \mathcal{G}_m \Lambda_{lz} \dots \Lambda_{lz} \mathcal{G}_1(\rho) \\ &= (\bar{\mathcal{G}}_m^\dagger \mathcal{Z}^{\otimes n} \bar{\mathcal{G}}_m) (\bar{\mathcal{G}}_{m-1}^\dagger \mathcal{Z}^{\otimes n} \bar{\mathcal{G}}_{m-1}) \dots (\bar{\mathcal{G}}_1^\dagger \mathcal{Z}^{\otimes n} \bar{\mathcal{G}}_1)(\rho) \\ &= \mathcal{P}_m \mathcal{P}_{m-1} \dots \mathcal{P}_1(\rho) \\ &= \mathcal{P}'(\rho),\end{aligned}\tag{89}$$

where  $\bar{\mathcal{G}}_i = \mathcal{G}_i \mathcal{G}_{i-1} \dots \mathcal{G}_1$ , and  $\mathcal{P}_i(\rho) = P \rho P^\dagger$  for non-identity  $n$ -qubit Pauli operator  $P \in \mathbf{P}_n \setminus \{\mathbb{I}\}$ . This implies that each circuit  $\vec{G}_i$  corresponds to a random Pauli operator  $\mathcal{P}'$  under the special local unitary noise channel  $\Lambda_{lz}(\theta = \frac{\pi}{2})$ . As a result, the value of  $\text{Tr}(Q \rho_f(\vec{G}_i))$  is either 0 or 1, then

$$\begin{aligned}\text{Tr}(Q \rho_f) &= \text{Tr}(Q \rho_f)^2, \\ A = \mathbb{E}(\text{Tr}(Q \rho_f)) &= \mathbb{E}(\text{Tr}(Q \rho_f)^2) = B,\end{aligned}\tag{90}$$

the reusing times  $R^*$  has a value equal to 0. On the other side, when  $\theta \rightarrow 0$ , the noise channel is close to identity channel  $\mathcal{I}$ ,  $B \approx A^2 \approx 1$ , so we have  $R^* \rightarrow \infty$ .

## C.3 Two-qubit correlation noise channel

As shown in Fig. 6, we investigate a 2-qubit composite channel

$$\Lambda_{comp3} = \Lambda_c(\beta_{12}) \circ \Lambda_{lp}(p_2) \circ \Lambda_{la}(p_1).\tag{91}$$

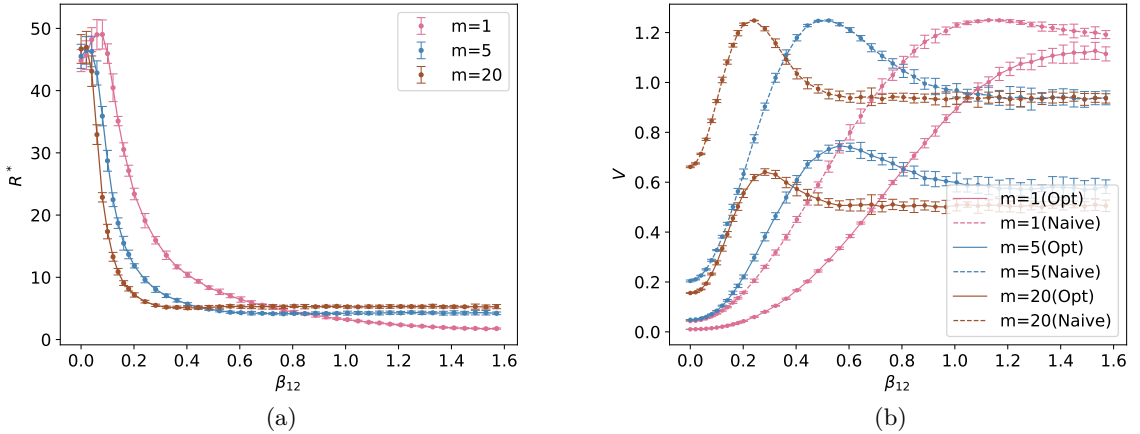


Figure 6: Figure (a) presents the optimal reusing times  $R^*$  for the 2-qubit standard RB protocol over composite noise channel  $\Lambda_{comp3}$  defined by Eq. (91). Figure (b) presents a comparison between the variance under the optimal  $R^*$  and that under the naive  $R_{naive} = 1$ . We omit the performance of the 2-optimal strategy  $R_0 = 4$  in the figures due to its variance being nearly identical to that of  $R^*$ , which would otherwise congest the visualization. The results under the  $\Lambda_{comp3}$  channel exhibit symmetry with respect to  $\beta_{12} = \frac{\pi}{2}$ ; hence, only the range of  $[0, \frac{\pi}{2}]$  for  $\beta_{12}$  is depicted in the figure.

The parameter  $\beta_{12}$  controls the noise strength of the interaction between two subsystems. The amplitude-damping and phase-damping parameters are assigned as  $p_1 = 0.999$



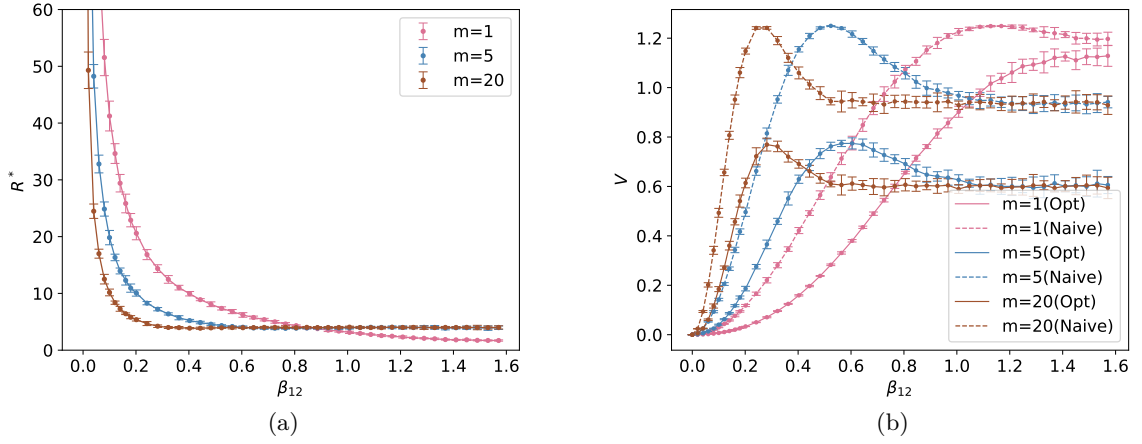


Figure 7: Figure (a) illustrates the values of the optimal reusing times  $R^*$  under the correlation channel  $\Lambda_c$ . Figure (b) presents a comparison between the variance under the optimal  $R^*$  and that under the naive  $R_{naive} = 1$ .

and  $p_2 = 0.99$ , respectively. When the noise parameter  $\beta_{12}$  increases, the value of  $R^*$  decreases. This occurs because as  $\beta_{12}$  approaches  $\frac{\pi}{2}$ , the noise induced by  $\Lambda_c(\beta_{12})$  intensifies, necessitating more twirling to learn about the noise channel. Note that in Fig. 6,  $R^*$  first slightly increases and then decreases as  $\beta_{12}$  approaching  $\frac{\pi}{2}$ . The increase results from the influence of another two types of composite noises  $\Lambda_{la}$  and  $\Lambda_{lp}$ . In the case where the noise channel is exclusively the correlation channel  $\Lambda_c$ ,  $R^*$  will exhibit a monotonic decrease as  $\beta_{12}$  increases, as illustrated in Fig. 7.

## References

- [1] A. W. Cross, G. Smith, and J. A. Smolin, Phys. Rev. A **92**, 012327 (2015), URL <https://doi.org/10.1103/PhysRevA.92.012327>.
- [2] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, et al., Science **376**, 1182 (2022), <https://www.science.org/doi/pdf/10.1126/science.abn7293>, URL <https://doi.org/10.1126/science.abn7293>.
- [3] H.-Y. Huang, R. Kueng, and J. Preskill, Nature Physics **16**, 1050 (2020), ISSN 1745-2481, URL <https://doi.org/10.1038/s41567-020-0932-7>.
- [4] M. Paris and J. Rehacek, *Quantum state estimation*, vol. 649 (Springer Science & Business Media, 2004), URL <https://doi.org/10.1007/b98673>.
- [5] R. B. A. Adamson and A. M. Steinberg, Phys. Rev. Lett. **105**, 030406 (2010), URL <https://doi.org/10.1103/PhysRevLett.105.030406>.
- [6] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2018), STOC 2018, p. 325–338, ISBN 9781450355599, URL <https://doi.org/10.1145/3188745.3188802>.
- [7] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, Phys. Rev. Lett. **112**, 190501 (2014), URL <https://doi.org/10.1103/PhysRevLett.112.190501>.
- [8] S. T. Flammia and J. J. Wallman, ACM Transactions on Quantum Computing **1** (2020), URL <https://doi.org/10.1145/3408039>.

- [9] S. T. Flammia and R. O'Donnell, *Quantum* **5**, 549 (2021), ISSN 2521-327X, URL <https://doi.org/10.22331/q-2021-09-23-549>.
- [10] J.-D. Urbina and K. Richter, *Advances in Physics* **62**, 363 (2013), URL <https://doi.org/10.1080/00018732.2013.860277>.
- [11] J. Emerson, R. Alicki, and K. Życzkowski, *Journal of Optics B: Quantum and Semi-classical Optics* **7**, S347 (2005), URL <https://doi.org/10.1088/1464-4266/7/10/021>.
- [12] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Phys. Rev. A* **77**, 012307 (2008), URL <https://doi.org/10.1103/PhysRevA.77.012307>.
- [13] H.-Y. Huang, R. Kueng, and J. Preskill, *Nature Physics* **16**, 1050 (2020), ISSN 1745-2481, URL <https://doi.org/10.1038/s41567-020-0932-7>.
- [14] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, *Phys. Rev. Lett.* **120**, 050406 (2018), URL <https://doi.org/10.1103/PhysRevLett.120.050406>.
- [15] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, *Science* **364**, 260 (2019), <https://www.science.org/doi/pdf/10.1126/science.aau4963>, URL <https://doi.org/10.1126/science.aau4963>.
- [16] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, *Nature Reviews Physics* **5**, 9 (2023), ISSN 2522-5820, URL <https://doi.org/10.1038/s42254-022-00535-2>.
- [17] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, *Nature* **603**, 416 (2022), ISSN 1476-4687, URL <https://doi.org/10.1038/s41586-021-04351-z>.
- [18] J. J. Wallman and S. T. Flammia, *New Journal of Physics* **16**, 103032 (2014), URL <https://doi.org/10.1088/1367-2630/16/10/103032>.
- [19] J. Helsen and M. Walter, *Phys. Rev. Lett.* **131**, 240602 (2023), URL <https://doi.org/10.1103/PhysRevLett.131.240602>.
- [20] Y. Zhou and Q. Liu, *Quantum* **7**, 1044 (2023), ISSN 2521-327X, URL <https://doi.org/10.22331/q-2023-06-29-1044>.
- [21] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. A* **85**, 042311 (2012), URL <https://doi.org/10.1103/PhysRevA.85.042311>.
- [22] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, *Phys. Rev. Lett.* **107**, 210404 (2011), URL <https://doi.org/10.1103/PhysRevLett.107.210404>.
- [23] I. L. Chuang and M. A. Nielsen, *Journal of Modern Optics* **44**, 2455 (1997), <https://www.tandfonline.com/doi/pdf/10.1080/09500349708231894>, URL <https://doi.org/10.1080/09500349708231894>.
- [24] J. F. Poyatos, J. I. Cirac, and P. Zoller, *Phys. Rev. Lett.* **78**, 390 (1997), URL <https://doi.org/10.1103/PhysRevLett.78.390>.
- [25] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, *Rev. Mod. Phys.* **95**, 045005 (2023), URL <https://doi.org/10.1103/RevModPhys.95.045005>.
- [26] O. Gühne and G. Tóth, *Physics Reports* **474**, 1 (2009), ISSN 0370-1573, URL <https://doi.org/10.1016/j.physrep.2009.02.004>.
- [27] S. Li, D. Fan, M. Gong, Y. Ye, X. Chen, Y. Wu, H. Guan, H. Deng, H. Rong, H.-L. Huang, et al., *Chin. Phys. Lett.* **39**, 030302 (2022), URL <http://doi.org/10.1088/0256-307X/39/3/030302>.
- [28] C. Dankert, R. Cleve, J. Emerson, and E. Livine, *Phys. Rev. A* **80**, 012304 (2009), URL <https://doi.org/10.1103/PhysRevA.80.012304>.

- [29] J. Helsen, I. Roth, E. Onorati, A. Werner, and J. Eisert, PRX Quantum **3**, 020357 (2022), URL <https://doi.org/10.1103/PRXQuantum.3.020357>.
- [30] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, Phys. Rev. A **89**, 062321 (2014), URL <https://doi.org/10.1103/PhysRevA.89.062321>.
- [31] J. Helsen, J. J. Wallman, and S. Wehner, Journal of Mathematical Physics **59**, 072201 (2018), ISSN 0022-2488, [https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.4997688/15970851/072201\\_1\\_online.pdf](https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.4997688/15970851/072201_1_online.pdf), URL <https://doi.org/10.1063/1.4997688>.
- [32] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, Phys. Rev. A **100**, 032304 (2019), URL <https://doi.org/10.1103/PhysRevA.100.032304>.
- [33] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, New Journal of Physics **17**, 113020 (2015), URL <https://doi.org/10.1088/1367-2630/17/11/113020>.
- [34] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010), URL <https://doi.org/10.1017/CB09780511976667>.
- [35] A. Streltsov, U. Singh, H. S. Dhar, M. N. Bera, and G. Adesso, Phys. Rev. Lett. **115**, 020403 (2015), URL <https://doi.org/10.1103/PhysRevLett.115.020403>.
- [36] M. Heinrich, M. Kliesch, and I. Roth, *Randomized benchmarking with random quantum circuits* (2023), **2212.06181**, URL <https://doi.org/10.48550/arXiv.2212.06181>.