

Markov chains for error accumulation in quantum circuits

Long Ma* and Jaron Sanders†

*Faculty of Electrical Engineering, Mathematics, and Computer Science
Delft University of Technology, PO Box 5031, 2600 GA Delft, the Netherlands*

We study a model for the accumulation of errors in multi-qubit quantum computations, as well as a model describing continuous errors accumulating in a single qubit. By modeling the error process in a quantum computation using two coupled Markov chains, we are able to capture a weak form of time-dependency between errors in the past and future. By subsequently using techniques from the field of discrete probability theory, we calculate the probability that error measures such as the fidelity and trace distance exceed a threshold analytically. The formulae cover fairly generic error distributions, cover multi-qubit scenarios, and are applicable to e.g. the randomized benchmarking protocol. To combat the numerical challenge that may occur when evaluating our expressions, we additionally provide an analytical bound on the error probabilities that is of lower numerical complexity, and we also discuss a state space reduction that occurs for stabilizer circuits. Finally, taking inspiration from the field of operations research, we illustrate how our expressions can be used to e.g. decide how many gates one can apply before too many errors accumulate with high probability, and how one can lower the rate of error accumulation in existing circuits through simulated annealing.

I. INTRODUCTION

The development of a quantum computer is expected to revolutionize computing by being able to solve hard computational problems faster than any classical computer [1]. However, present-day state-of-the-art quantum computers are prone to errors in their calculations due to physical effects such as unwanted qubit–qubit interactions, qubit crosstalk, and state leakage [2]. Minor errors can be corrected, but error correction methods will still be overwhelmed once too many errors occur [3–5]. Quantum circuits with different numbers of qubits and circuit depths have been designed to implement algorithms more reliably [6], and the susceptibility of a circuit to the accumulation of errors is an important evaluation criterion. We therefore study now Markov chains that describe the accumulation of errors in quantum circuits. Different types of errors [7] that can occur and are included in our model are e.g. Pauli channels [1], Clifford channels [8, 9], depolarizing channels [1], and small rotational errors [10, 11]. If the random occurrence of such errors only depends on the last state of the quantum mechanical system, the probability that error measures such as the fidelity and trace distance accumulate beyond a threshold can be related to different hitting time distributions of two coupled Markov chains. These hitting time distributions are then calculated analytically using techniques from probability theory and operations research.

Error accumulation models that share similarities with the Markov chains under consideration here can primarily be found in the literature on randomized benchmarking [12]. From the modeling point of view, the dynamical description of error accumulation that we adopt is shared in [13–16]. These articles however do not explicitly tie

the statistics of error accumulation to a hitting time analysis of a coupled Markov chain. Furthermore, while Markovianity assumptions on e.g. noise are common [17], the explicit mention of an underlying random walk is restricted to a few papers only [14, 18]. From the analysis point of view, research on randomized benchmarking has predominantly focused on generalizing expressions for the expected fidelity over time. For example, the expected decay rates of the fidelity are analyzed for cases of randomized benchmarking with restricted gate sets [19], Gaussian noise with time-correlations [20], gate-dependent noise [16], and leakage errors [21]; and the expected loss rate of a protocol related to randomized benchmarking is calculated in [22]. In this article, we focus instead on the probability distributions of both the error and maximum error – which capture the statistics in more detail than an expectation – for arbitrary distance measures, and in random as well as nonrandom quantum circuits. Finally, [13, 14, 16, 21] resort to perturbation or approximate analyses (via e.g. Taylor expansions, and independence or decorrelation assumptions) to characterize the fidelity, whereas here we provide the exact, closed-form expressions for the distributions. Our bridging of techniques from probability theory and operations research to the domain of quantum computing is a new angle.

To be precise: this article first studies a model for discrete Markovian error accumulation in a multi-qubit quantum circuit. We suppose for simplicity that both the quantum gates and errors belong to a finite unitary group $\mathcal{G}_n \subseteq \mathcal{U}(2^n)$, where $\mathcal{U}(2^n)$ is the unitary group for n qubits. The group \mathcal{G}_n can e.g. be the generalized Pauli group (i.e., the discrete Heisenberg–Weyl group), or the Clifford group. By modeling the quantum computation with and without errors as two coupled Markov chains living on the state space consisting of pairs of elements from these groups, we are able to capture a weak form of time-dependency within the process of error accumulation. To see this, critically note that the assumption of a Markov property does not imply that the past and the future in

* l.ma-2@tudelft.nl

† Corresponding author: jaron.sanders@tue.nl

the quantum computation are independent given any information concerning the present [23]. We must also note that while the individual elements of our two-dimensional Markov chain belong to a group, the two-dimensional Markov chain itself, here, is generally not a random walk on a group. Lastly, our Markov chain model works for an arbitrary number of qubits. These model features are all relevant to the topic of error modeling in quantum computing, and since the Markov property is satisfied in randomized benchmarking, the model has immediate application. Secondly, in this article we briefly study a known random walk model on the three-dimensional sphere [24]. This model is commonly used to describe the average dephasing of a single qubit (or spin) [25]. We characterize the distribution and expectation of the trace distance measuring the error that is accumulated over time. This is, essentially, a refinement to provide information about the higher-order statistics of the error accumulation in a single qubit.

It should be noted that the numerical complexity of our exact expressions, however, can be high for large quantum circuits. The precise difficulty of evaluating our expressions depends on the particulars of the quantum circuit one looks at. For practical purposes, we therefore also provide an analytical bound on the maximum error probability that is of lower numerical complexity. Additionally, we discuss the reduction in complexity that occurs when starting a quantum computation from a stabilizer state: the coupled Markov chain's state space then reduces in size. The article [15] also notes the high computational complexity of error analysis in quantum circuits in general. The issue is there approached combinatorially by converting circuits into directed graphs, tracing so-called fault-paths through these graphs, and therewith calculating or estimating the success rates of circuits.

Finally, we use the expressions that describe how likely it is that errors accumulate to answer two operational questions that will help advance the domain of practical quantum computing [26]. First, we calculate and bound analytically how many quantum gates $t_{\delta,\gamma}^*$ one can apply before an error measure of your choice exceeds a threshold δ with a probability above γ . This information is useful for e.g. deciding how often a quantum computer should perform repairs on qubits, and is particularly opportune at this moment since quantum gates fail $O(0.1\text{--}1\%)$ of the time [26]. Related but different ideas can be found in e.g. [7, §2.3], where the accumulation of bit-flips and rotations on a repetition code is studied and a time to failure is derived, and in [27, §V], where an upper bound on the number of necessary measurements for a randomized benchmarking protocol is derived. Second, using techniques from optimization, we design a simulated annealing method that improves existing circuits by swapping out gate pairs to achieve lower rates of error accumulation. We also discuss conditions under which this tailor-made method is guaranteed to find the best possible circuit. Both of these excursions illustrate how the availability of

an analytical expression for the accumulation of errors allows us to proceed with second-tier optimization methods to facilitate quantum computers in the long-term.

This article is structured as follows. In Section II, we give the model aspects pertaining to the quantum computation (gates, error dynamics, and error measures) and we introduce the coupled Markov chain that describes error accumulation. In Section III, we provide the relation between the probability of error and the hitting time distributions, and we derive the error distributions as well as its bound. We also calculate the higher-order statistics of an error accumulation model for a single qubit that undergoes (continuous) random phase kicks and depolarization. In Section IV, we illustrate our theoretical results by comparing to numerical results of a quantum simulator we wrote for this article. In Section V, we discuss the simulated annealing scheme. Finally, in Section VI, we conclude with ideas for future research.

II. MODEL AND COUPLED MARKOV CHAIN

A. Gates and errors in quantum computing

It is generally difficult to describe large quantum systems on a classical computer for the reason that the state space required increases exponentially in size with the number of qubits [28]. However, the stabilizer formalism is an efficient tool to analyze such complex systems [29]. Moreover, the stabilizer formalism covers many paradoxes in quantum mechanics [30], including the Greenberger–Horne–Zeilinger (GHZ) experiment [31], dense quantum coding [32], and quantum teleportation [33]. Specifically, the stabilizer circuits are the smallest class of quantum circuits that consist of the following four gates: $\omega = e^{i\pi/4}$, $H = (1/\sqrt{2}) \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, $S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$, and

$$Z_c = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

These four gates are closed under the operations of tensor product and composition [34]. As a consequence of the Gottesman–Knill theorem, stabilizer circuits can be efficiently simulated on a classical computer [35].

Unitary stabilizer circuits are also known as the Clifford circuits; the Clifford group \mathcal{C}_n can be defined as follows. First: let $P \triangleq \{I, X, Y, Z\}$ denote the Pauli matrices,¹ and let $P_n \triangleq \{\sigma_1 \otimes \cdots \otimes \sigma_n \mid \sigma_i \in P\}$ denote the Pauli matrices on n qubits. The Pauli matrices are commonly used to model errors that can occur due to the interactions of the qubit with its environment [36]. In the case of a single qubit, the matrix I represents that there is no error, the matrix X that there is a bit-flip error, the matrix Z that there is a phase-flip error, and the matrix Y that there are both a bit-flip and a phase-flip error.

¹ Here, $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, and $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

The multi-qubit case interpretations follows analogously. Second: let $P_n^* = P_n \setminus I^{\otimes n}$. We now define the Clifford group on n qubits by $\mathcal{C}_n \triangleq \{U \in \mathcal{U}(2^n) \mid \sigma \in \pm P_n^* \Rightarrow U\sigma U^\dagger \in \pm P_n^*\} \setminus \mathcal{U}(1)$. The fact that \mathcal{C}_n is a group can be verified by checking the two necessary properties (see Appendix A). The Clifford group on n qubits is finite [37] and is of size

$$|\mathcal{C}_n| = 2^{n^2+2n} \prod_{i=1}^n (4^i - 1).$$

Here we ignore the global phase since it has no physical significance. For a single qubit, a representation for the Clifford group $\mathcal{C}_1 = \{c_1, c_2, \dots, c_{24}\}$ can be enumerated and its elements are for example shown in [12] and [14].

B. Dynamics of error accumulation

Suppose that we had a faultless, perfect quantum computer. Then a faultless quantum mechanical state ρ_t at time t could be calculated under a gate sequence $\mathcal{U}_t = \{U_1, \dots, U_t\}$ from the initial state $\rho_0 \triangleq |\psi_0\rangle\langle\psi_0|$. Here $\tau < \infty$ denotes the sequence length, and $t \in \{0, 1, \dots, \tau\}$ enumerates the intermediate steps. On the other hand, with an imperfect quantum computer, a possibly faulty quantum mechanical state σ_t at time t would be calculated under both \mathcal{U}_t and some (unknown) noise sequence $\mathcal{E}_t = \{\Lambda_1, \dots, \Lambda_t\}$ starting from an initial state $\sigma_0 \triangleq |\Psi_0\rangle\langle\Psi_0|$ possibly different from ρ_0 . We define the set of all pure states for n qubits as \mathcal{S}^n and consider the situation that $|\psi_0\rangle, |\Psi_0\rangle \in \mathcal{S}^n$.

To be precise, define for the faultless quantum computation

$$\rho_t \triangleq |\psi_t\rangle\langle\psi_t| = U_t |\psi_{t-1}\rangle\langle\psi_{t-1}| U_t^\dagger$$

for times $t = 1, 2, \dots, \tau$. Let $X_t \triangleq U_t U_{t-1} \dots U_1$ be shorthand notation such that $\rho_t = X_t \rho_0 X_t^\dagger$. For the possibly faulty quantum computation, define

$$\sigma_t \triangleq |\Psi_t\rangle\langle\Psi_t| = \Lambda_t U_t |\Psi_{t-1}\rangle\langle\Psi_{t-1}| U_t^\dagger \Lambda_t^\dagger$$

for times $t = 1, 2, \dots, \tau$, respectively. Introduce also the shorthand notation $Y_t \triangleq \Lambda_t U_t \Lambda_{t-1} U_{t-1} \dots \Lambda_1 U_1$ such that $\sigma_t = Y_t \sigma_0 Y_t^\dagger$. The analysis in this paper can immediately be extended to the case where errors (also) precede the gate. The error accumulation process is also illustrated in Figure 1.

C. Distance measures for quantum errors

The error can be quantified by any measure of distance between the faultless quantum-mechanical state ρ_t and the possibly faulty quantum-mechanical state σ_t for steps $t = 0, 1, \dots, \tau$. For example, we can use the fidelity $F_t \triangleq \text{Tr} \sqrt{\rho_t^{1/2} \sigma_t \rho_t^{1/2}}$ [1], or the Schatten p -norm [38]

a) Faultless computation:

$$\rho_0 \xrightarrow{U_1} \rho_1 \xrightarrow{U_2} \dots \xrightarrow{U_{\tau-1}} \rho_{\tau-1} \xrightarrow{U_\tau} \rho_\tau$$

b) Potentially faulty computation:

$$\sigma_0 \xrightarrow{\Lambda_1 U_1} \sigma_1 \xrightarrow{\Lambda_2 U_2} \dots \xrightarrow{\Lambda_{\tau-1} U_{\tau-1}} \sigma_{\tau-1} \xrightarrow{\Lambda_\tau U_\tau} \sigma_\tau$$

FIG. 1. Schematic depiction of the coupled quantum mechanical states ρ_t and σ_t for times $t = 0, 1, \dots, \tau$. a) Faultless computation. The state ρ_t is calculated based on a gate sequence $\mathcal{U}_t = \{U_1, \dots, U_t\}$ from the initial state ρ_0 . b) Potentially faulty computation. The state σ_t is calculated using *the same* gate sequence $\mathcal{U}_t = \{U_1, \dots, U_t\}$ and an additional error sequence $\mathcal{E}_t = \{\Lambda_1, \dots, \Lambda_t\}$. The final state σ_τ can depart from the faultless state ρ_τ because of errors.

defined by

$$D_t \triangleq \|\sigma_t - \rho_t\|_p = \frac{1}{2} \text{Tr} \left[\left\{ (\sigma_t - \rho_t)^\dagger (\sigma_t - \rho_t) \right\}^{\frac{p}{2}} \right]^{\frac{1}{p}}$$

for any $p \in [1, \infty)$. The Schatten p -norm reduces to the trace distance for $p = 1$, the Frobenius norm for $p = 2$, and the spectral norm for $p = \infty$. In the case of one qubit, the trace distance between quantum-mechanical states ρ_t and σ_t equals half of the Euclidean distance between ρ_t and σ_t when representing them on the Bloch sphere [1]. It is well known that the trace distance is invariant under unitary transformations [1]; a fact that we leverage in Section III.

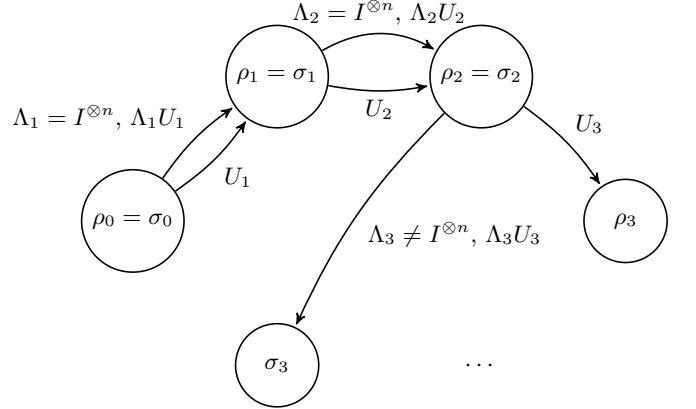


FIG. 2. Coupled chain describing the quantum circuit with errors. In this depiction, we start from *the same* initial state for simplicity. Here an error $\Lambda_3 \neq I^{\otimes n}$ occurs as the third gate is applied. Note that the coupled chain ρ_t, σ_t separates.

In this paper, we are going to analyze the statistical properties of some arbitrary distance measure (one may choose) between the quantum mechanical states ρ_t and σ_t for times $t = 0, 1, \dots, \tau$. For illustration, we will state the results in terms of the Schatten p -norm, and so are

after its expectation $\mathbb{E}[D_t]$, as well as its distributions $\mathbb{P}[D_t \leq \delta]$, $\mathbb{P}[\max_{0 \leq s \leq t} D_s \leq \delta]$. As we show in §B, in case of the trace distance ($p = 1$), these probabilities can then be related to the corresponding probabilities for the fidelity:

Lemma 1. *With $p = 1$, it holds that $\mathbb{P}[F_t \geq 1 - \varepsilon] \geq \mathbb{P}[D_t \leq \varepsilon]$ for all $t \geq 0$. Furthermore,*

$$\mathbb{P}[\min_{0 \leq s \leq t} F_s \geq 1 - \varepsilon] \geq \mathbb{P}[\max_{0 \leq s \leq t} D_s \leq \varepsilon].$$

III. ERROR ACCUMULATION

A. Discrete, random error accumulation (multi-qubit case)

Following the model described in Section II and illustrated in Figure 1 and Figure 2, we define the gate pairs $Z_t \triangleq (X_t, Y_t)$ for $t = 1, 2, \dots, \tau$, and suppose that $Z_0 = z_0$ is given a priori. Note that if one assumes $\rho_0 = \sigma_0$, then $z_0 = (I^{\otimes n}, I^{\otimes n})$.

1. The case of random circuits

We consider first the scenario that each next gate is selected randomly and independently from everything but the last system state. This assumption is satisfied in e.g. the randomized benchmarking protocol [12–22]. The probabilities $\mathbb{P}_{z_0}[D_t > \delta]$ and $\mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s \leq \delta]$ can then be calculated once the initial states $|\psi_0\rangle$, $|\Psi_0\rangle$ and the *transition matrix* are known.

Let the transition matrix of the Markov chain $\{Z_t\}_{t \geq 0}$ be denoted element-wise by $P_{z,w} \triangleq \mathbb{P}[Z_{t+1} = w | Z_t = z]$ for $z = (x, y), w = (u, v) \in \mathcal{G}_n^2$. The transition matrix satisfies $P \in [0, 1]^{|\mathcal{G}_n|^2 \times |\mathcal{G}_n|^2}$ and the elements of each of its rows sum to one. Let

$$P_{z_0, w}^{(t)} \triangleq \mathbb{P}[Z_t = w | Z_0 = z_0] = (P^t)_{z_0, w} \quad (1)$$

stand in for the probability that the process is at state w at time t starting from $Z_0 = z_0$. Note that the second equality follows from the Markov property [23].

Example 1: Consider the situation that the error depends on the last gate. The transition probability $P_{z,w}$ for $z = (x, y), w = (u, v) \in \mathcal{G}_n^2$ can then be calculated as follows. For the faultless computation, a gate $U = ux^{-1}$ that transfers the density matrix $x\rho_0x^\dagger$ to $u\rho_0u^\dagger$ is randomly chosen according to a gate probability vector κ . For the possibly faulty computation, an error that transfers the density matrix $y\sigma_0y^\dagger$ to $v\sigma_0v^\dagger$, after the gate $U = ux^{-1}$, is $\Lambda = vy^{-1}xu^{-1}$. Let $\zeta(\Lambda = vy^{-1}xu^{-1} | ux^{-1})$ denote the probability that the error $\Lambda = vy^{-1}xu^{-1}$ occurs given that the gate $U = ux^{-1}$ just occurred. The transition matrix then satisfies $\mathbb{P}[Z_{t+1} = w | Z_t = z] = \kappa(U = ux^{-1})\zeta(\Lambda = vy^{-1}xu^{-1} | ux^{-1})$ component-wise.

Example 2: If we assume that errors and gates are independently generated, then the transition matrix satisfies $\mathbb{P}[Z_{t+1} = w | Z_t = z] = \kappa(U = ux^{-1})\zeta(\Lambda = vy^{-1}xu^{-1})$ component-wise.

We are now after the probability that the distance D_t is larger than a threshold δ . We define thereto the set of δ -bad gate pairs by

$$\mathcal{B}_{|\psi_0\rangle, \delta}^{\Psi_0} \triangleq \{(x, y) \in \mathcal{G}_n^2 \mid \|x\rho_0x^\dagger - y\sigma_0y^\dagger\|_p > \delta\} \quad (2)$$

for $|\psi_0\rangle, |\Psi_0\rangle \in \mathcal{S}^n, \delta \geq 0$, as well as the *hitting time* of any set $\mathcal{A} \subseteq \mathcal{G}_n^2$ by

$$T_{\mathcal{A}} \triangleq \inf\{t \geq 0 \mid Z_t \in \mathcal{A}\} \quad (3)$$

with the convention that $\inf \emptyset = \infty$. Note that $T_{\mathcal{A}} \in \mathbb{N}_0 \cup \{\infty\}$ and that it is random. With definitions (2), (3), we have the convenient representation

$$\begin{aligned} \mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s \leq \delta] &= 1 - \mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s > \delta] \\ &= 1 - \mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{\Psi_0}} \leq t] \end{aligned} \quad (4)$$

for this homogeneous Markov chain. As a consequence of (4), the analysis comes down to an analysis of the hitting time distribution for this coupled Markov chain (Figure 3).

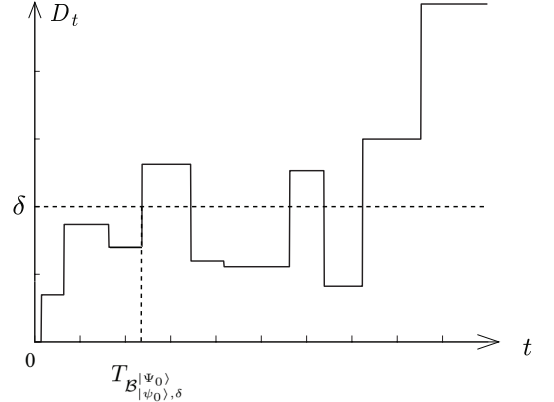


FIG. 3. Schematic diagram of the hitting time $T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{\Psi_0}}$.

Define the matrix $B_{|\psi_0\rangle, \delta}^{\Psi_0} \in [0, 1]^{|\mathcal{G}_n|^2 \times |\mathcal{G}_n|^2}$ element-wise by

$$(B_{|\psi_0\rangle, \delta}^{\Psi_0})_{z,w} \triangleq \begin{cases} P_{z,w} & \text{if } w \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{\Psi_0}, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Let the initial state vector be denoted by e_{z_0} , a $|\mathcal{G}_n|^2 \times 1$ vector with just the z_0 -th element 1 and the others 0. Also let $1_{\mathcal{A}}$ denote the $|\mathcal{G}_n|^2 \times 1$ vector with ones in every coordinate corresponding to an element in the set \mathcal{A} . Finally, we define a $|\mathcal{G}_n|^2 \times 1$ vector $d_{|\psi_0\rangle}^{\Psi_0} = (\|x\rho_0x^\dagger - y\sigma_0y^\dagger\|_p)_{(x,y) \in \mathcal{G}_n^2}$ enumerating all possible Schatten p -norm distances. We now state our first result:

Proposition 2 (Error accumulation in random circuits). *For any $z_0 \in \mathcal{G}_n^2$, $\delta \geq 0$, $t = 0, 1, \dots, \tau < \infty$: the distribution of error is given by*

$$\mathbb{P}_{z_0}[D_t > \delta] = e_{z_0}^T P^t 1_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}}, \quad (6)$$

and is nonincreasing in δ . Similarly, the expected error is given by $\mathbb{E}_{z_0}[D_t] = e_{z_0}^T P^t d_{|\psi_0\rangle}^{|\Psi_0\rangle}$. Furthermore; if $z_0 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}$, the distribution of maximum error is given by

$$\begin{aligned} \mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s > \delta] \\ = \sum_{s=1}^t e_{z_0}^T (B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle})^{s-1} (P - B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}) 1_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}}, \end{aligned} \quad (7)$$

and otherwise it equals one. Lastly, (7) is nonincreasing in δ , and nondecreasing in t .

The probability in (7) is a more stringent error measure than e.g. (6) is. The event $\{\max_{0 \leq s \leq t} D_s < \delta\}$ implies after all that the error D_t has always been below the threshold δ up to and including at time t . The expected error $\mathbb{E}_{z_0}[D_t]$ and distribution $\mathbb{P}_{z_0}[D_t > \delta]$ only concern the error at time t . Additionally, (7) allows us to calculate the maximum number of gates that can be performed. That is, $\mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s > \delta] \leq \gamma$ as long as

$$t \leq t_{\delta, \gamma}^* \triangleq \max_{s \in \mathbb{N}_+} \{t | \mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s > \delta] \leq \gamma\}. \quad (8)$$

In words: at most $t_{\delta, \gamma}^*$ gates can be applied before an accumulated error of size at least δ occurred with probability at least γ .

Proof of (6). It follows from (2), mutual exclusivity, and (1) that

$$\begin{aligned} \mathbb{P}_{z_0}[D_t > \delta] &= \mathbb{P}_{z_0}[Z_t \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}] \\ &= \sum_{w \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} \mathbb{P}_{z_0}[Z_t = w] = \sum_{w \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} (P^t)_{z_0, w}. \end{aligned}$$

The right-hand side equals (6) in matrix notation. To obtain the expression for the expectation, directly apply the definition of expectation for a discrete random variable:

$$\mathbb{E}_{z_0}[D_t] = \sum_{(x, y) \in \mathcal{G}_n^2} \|x \rho_0 x^\dagger - y \sigma_0 y^\dagger\|_p \mathbb{P}_{z_0}[Z_t = (x, y)].$$

Using (1) and the definition of $d_{|\psi_0\rangle}^{|\Psi_0\rangle}$, this gives the result.

Proof of (7). If $z_0 \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}$, then $\mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} = 0] = 1$. If $z_0 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}$, then use (5) to write

$$\begin{aligned} \mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} = s] \\ = \mathbb{P}_{z_0}[Z_1 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}, \dots, Z_{s-1} \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}, Z_s \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}] \end{aligned}$$

$$\begin{aligned} &= \sum_{z_1 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} \dots \sum_{z_{s-1} \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} \sum_{z_s \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} \mathbb{P}_{z_0}[Z_1 = z_1, \dots, Z_s = z_s] \\ &= e_{z_0}^T (B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle})^{s-1} (P - B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}) 1_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} \end{aligned} \quad (9)$$

in matrix notation. The result follows after summing (9) for $s = 0, 1, \dots, t-1$ by mutual exclusivity.

Note finally that for arbitrary $\delta_2 > \delta_1$, we have that $\mathcal{B}_{|\psi_0\rangle, \delta_2}^{|\Psi_0\rangle} \subseteq \mathcal{B}_{|\psi_0\rangle, \delta_1}^{|\Psi_0\rangle}$. As a consequence,

$$\mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta_2}^{|\Psi_0\rangle}} \leq t] \leq \mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta_1}^{|\Psi_0\rangle}} \leq t].$$

This establishes that $\mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} \leq t]$ is nonincreasing in δ . By positivity of the summands, $\mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} \leq t]$ is nondecreasing in t . \square

For general $\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}$, the explicit calculation of (7) can be numerically intensive. It is however possible to provide a lower bound of lower numerical complexity via the expected hitting time of the set $\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}$:

Lemma 3 (Lower bound for random circuits). *For any set $\mathcal{A} \subseteq \mathcal{G}_n^2$, the expected hitting times of a homogeneous Markov chain are the solutions to the linear system of equations $\mathbb{E}_z[T_{\mathcal{A}}] = 0$ for $z \in \mathcal{A}$, $\mathbb{E}_z[T_{\mathcal{A}}] = 1 + \sum_{w \notin \mathcal{A}} P_{z, w} \mathbb{E}_w[T_{\mathcal{A}}]$ for $z \notin \mathcal{A}$. Furthermore; for any $z_0 \in \mathcal{G}_n^2$, $\delta \geq 0$, $t = 0, 1, \dots, \tau < \infty$:*

$$\mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s > \delta] \geq 0 \vee \left(1 - \frac{\mathbb{E}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}}]}{t+1}\right). \quad (10)$$

Here $a \vee b \triangleq \max\{a, b\}$.

As a consequence of Lemma 3,

$$\mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s > \delta] \geq \gamma \text{ when } t \geq \frac{\mathbb{E}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}}]}{1 - \gamma} - 1,$$

and in particular

$$\mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s > 0] > 0 \text{ when } t \geq \mathbb{E}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, 0}^{|\Psi_0\rangle}}].$$

The values in the right-hand sides are thus upper bounds to the number of gates $t_{\delta, \gamma}^*$ one can apply before δ error has occurred with probability γ :

$$t_{\delta, \gamma}^* \leq \mathbb{E}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, 0}^{|\Psi_0\rangle}}] \wedge \left(\frac{\mathbb{E}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}}]}{1 - \gamma} - 1\right)$$

for $\delta \geq 0, \gamma \in [0, 1]$. Here, $a \wedge b \triangleq \min\{a, b\}$.

Proof of (10). The first part is a standard result, see e.g. [39, p. 202]. The second part follows from Markov's inequality, i.e.,

$$\mathbb{P}_{z_0}[\max_{0 \leq s \leq t} D_s \leq \delta] = \mathbb{P}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}} > t] \leq \frac{\mathbb{E}_{z_0}[T_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}}]}{t+1}.$$

That is it. \square

2. The case of nonrandom circuits

Suppose that the gate sequence $\mathcal{U}_\tau = \{U_1, \dots, U_\tau\}$ is fixed *a priori* and that it is not generated randomly. Because the gate sequence is nonrandom, we have now that the faultless state $\rho_t = X_t \rho_0 X_t^\dagger$ is deterministic for times $t = 0, 1, \dots, \tau$. On the other hand the potentially faulty state $\sigma_t = Y_t \rho_0 Y_t^\dagger$ is still (possibly) random.

We can now use a lower dimensional Markov chain to represent the system. To be precise: we will now describe the process $\{\sigma_t\}_{t \geq 0}$ as an *inhomogeneous Markov chain*. Its transition matrices will now be time-dependent and given element-wise by $Q_{y,v}(t) = \mathbb{P}[\sigma_{t+1} = v \rho_0 v^\dagger | \sigma_t = y \rho_0 y^\dagger]$ for $y, v \in \mathcal{G}_n$, $t \in \{0, 1, \dots, \tau - 1\}$. Letting $Q_{y,v}^{(t)} \triangleq \mathbb{P}[\sigma_t = v \rho_0 v^\dagger | \sigma_0 = y \rho_0 y^\dagger]$ stand in for the probability that the process is at state $v \rho_0 v^\dagger$ at time t starting from $\sigma_0 = y \rho_0 y^\dagger$, we have by the Markov property [23] that

$$Q_{y,v}^{(t)} = \left(\prod_{s=1}^t Q(s) \right)_{y,v} \quad \text{for } y, v \in \mathcal{G}_n. \quad (11)$$

Example 3: Consider the situation that the probability that an error occurs depends on which gate was applied last. If we assume that $\mathbb{P}[\Lambda_{t+1} = \lambda | Y_t = y] = \zeta_{y, U_{t+1}}(\lambda)$ are given distributions for $y \in \mathcal{G}_n$, $t \in \{0, 1, \dots, \tau - 1\}$ on $\lambda \in \mathcal{G}_n$, we can alternatively write the elements of the transition matrices as

$$\begin{aligned} Q_{y,v}(t) &= \mathbb{P}[Y_{t+1} = v | Y_t = y] \\ &= \sum_{\lambda \in \mathcal{G}_n} \mathbb{P}[Y_{t+1} = v | Y_t = y, \Lambda_{t+1} = \lambda] \mathbb{P}[\Lambda_{t+1} = \lambda | Y_t = y] \\ &= \sum_{\lambda \in \mathcal{G}_n} \mathbb{1}[\lambda U_{t+1} y \rho_0 y^\dagger U_{t+1}^\dagger \lambda^\dagger = v \rho_0 v^\dagger] \zeta_{y, U_{t+1}}(\lambda). \end{aligned}$$

Here, we have used the law of total probability.

Example 4: If errors occur independently and with probability $\mathbb{P}[\Lambda_{t+1} = \lambda] = \zeta(\lambda)$, then $Q_{y,v}(t) = \sum_{\lambda \in \mathcal{G}_n} \mathbb{1}[\lambda U_{t+1} y \rho_0 y^\dagger U_{t+1}^\dagger \lambda^\dagger = v \rho_0 v^\dagger] \zeta(\lambda)$.

Now define the sets of (δ, t) -bad gate pairs by $\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t} \triangleq \{x \in \mathcal{U}_n | \|\rho_t - x \sigma_0 x^\dagger\|_p > \delta\}$ for $|\psi_0\rangle, |\Psi_0\rangle \in \mathcal{S}^n$, $t \in \{0, 1, \dots, \tau\}$, $\delta \geq 0$. Also define the matrices $B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t} \in [0, 1]^{|\mathcal{G}_n| \times |\mathcal{G}_n|}$ element-wise by

$$(B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t})_{y,v} \triangleq \begin{cases} Q_{y,v}(t) & \text{if } v \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t}, \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

for $t = 0, 1, \dots, \tau$. Recall the notation introduced above Proposition 2. Similarly enumerate in the vector d_{ρ_t} the Schatten p -norms between any of the possible states of σ_t and the faultless state ρ_t . We state our second result:

Proposition 4 (Error accumulation in nonrandom circuits). *For any $y_0 \in \mathcal{G}_n$, $\delta \geq 0$, $t = 0, 1, \dots, \tau < \infty$: the*

distribution of error is given by

$$\mathbb{P}_{y_0}[D_t > \delta] = e_{y_0}^T \left(\prod_{k=1}^t Q(k) \right) \mathbf{1}_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t}}. \quad (13)$$

Similarly, the expected error is given by $\mathbb{E}_{y_0}[D_t] = e_{y_0}^T \left(\prod_{k=1}^t Q(k) \right) d_{\rho_t}$. Furthermore; if $y_0 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, 0}$, the distribution of maximum error is given by

$$\begin{aligned} \mathbb{P}_{y_0}[\max_{0 \leq s \leq t} D_s > \delta] &= \sum_{s=0}^{t-1} \left(e_{y_0}^T \left(\prod_{r=0}^s B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, r} \right) \right. \\ &\quad \times (Q(s+1) - B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s+1}) \mathbf{1}_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s+1}} \Big), \end{aligned} \quad (14)$$

and otherwise it equals one.

Proof of (13). From $\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t}$'s definition and mutual exclusivity it follows immediately that

$$\mathbb{P}_{y_0}[D_t > \delta] = \mathbb{P}_{y_0}[Y_t \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t}] = \sum_{v \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t}} \mathbb{P}_{y_0}[Y_t = v] \quad (15)$$

for $|\psi_0\rangle, |\Psi_0\rangle \in \mathcal{S}^n, \delta \geq 0$. Using (11) and continuing from (15), we obtain

$$\mathbb{P}_{y_0}[D_t > \delta] = \sum_{v \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t}} e_{y_0}^T \left(\prod_{k=1}^t Q(k) \right)_{y,v}.$$

This simplifies to (13) in matrix notation. To obtain the expression for the expectation, apply the same arguments as were used for Proposition 2, but use (11) instead.

Proof of (14). We can again explicitly calculate the result using a hitting time analysis, but the expressions expand due to the time-dependency of $\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, t}$. If $y_0 \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, 0}$, then $\mathbb{P}_{y_0}[\max_{0 \leq r \leq s} D_r > \delta] = 1$. Otherwise

$$\mathbb{P}_{y_0}[\{\max_{0 \leq r \leq s-1} D_r \leq \delta\} \cap \{D_s > \delta\}] \quad (16)$$

$$= \mathbb{P}_{y_0}[Y_1 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, 1}, \dots, Y_{s-1} \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s-1}, Y_s \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s}]$$

$$\begin{aligned} &= \sum_{y_1 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, 1}} \dots \sum_{y_s \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s}} \mathbb{P}_{y_0}[Y_1 = y_1, \dots, Y_s = y_s] \\ &= \sum_{y_1 \notin \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, 1}} \dots \sum_{y_{s-1} \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s-1}} \sum_{y_s \in \mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s}} \prod_{r=0}^{s-1} Q_{y_r, y_{r+1}}(r). \end{aligned}$$

Recalling (12), we can equivalently write (16) in matrix notation as

$$\begin{aligned} &\mathbb{P}_{y_0}[\{\max_{0 \leq r \leq s-1} D_r \leq \delta\} \cap \{D_s > \delta\}] \\ &= e_{y_0}^T \left(\prod_{r=1}^{s-1} B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, r} \right) (Q(s) - B_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s}) \mathbf{1}_{\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle, s}}. \end{aligned} \quad (17)$$

Summing (17) over $s = 0, 1, \dots, t-1$ completes the proof by mutual exclusivity. \square

3. State space reduction in stabilizer circuits

The set of stabilizer gates [40] for a state $|\psi\rangle$ is defined as the set of gates $\mathcal{M} \in \mathcal{G}_n \setminus I^{\otimes n}$ that satisfy $\mathcal{M}|\psi\rangle = e^{i\gamma}|\psi\rangle$ for some $\gamma \in \mathbb{R}$. Since $e^{i\gamma}$ is a global phase that cannot be observed, $\mathcal{M}|\psi\rangle = e^{i\gamma}|\psi\rangle$ can also be understood as part of an equivalence class $\mathcal{M}|\psi\rangle \equiv |\psi\rangle$. The state $|\psi\rangle$ in $\mathcal{M}|\psi\rangle \equiv |\psi\rangle$ is called the *stabilizer state* [41]. For one qubit and in case of the Pauli group, examples include $|0\rangle$, $|1\rangle$, and $|\pm\rangle = (1/2)(|0\rangle \pm |1\rangle)$. Remark 5 shows that there exist 2^n stabilizer states for any gate $\mathcal{M} \in \mathcal{G}_n \setminus I^{\otimes n}$. Its proof is relegated to §C.

Remark 5. For any gate $\mathcal{M} \in \mathcal{G}_n \setminus I^{\otimes n}$ there are 2^n states $|\psi_0\rangle$ that satisfy $\mathcal{M}|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$ for some $\gamma \in \mathbb{R}$.

The advantage of starting a quantum circuit from a stabilizer state is that the state space is smaller. It moreover can be proved that, under the assumptions of Section II, when starting initially from a stabilizer state, all states reached during the quantum computation will themselves be stabilizer states. Define the set of *reachable density matrices* from an initial state $|\psi_0\rangle \in \mathcal{S}^n$, by

$$\mathcal{R}_{|\psi_0\rangle} \triangleq \{g|\psi_0\rangle \mid g \in \mathcal{G}_n\}. \quad (18)$$

The exact number of reachable states can be calculated by the method in §F. Taking the Clifford group gates on two qubits as an example, the number of gates $|\mathcal{C}_2| = 11520$. However, there are just 60 reachable states if the initial state is $|00\rangle$. The proof of Remark 6 can be found in §D.

Remark 6. Given a gate $\mathcal{M} \in \mathcal{G}_n \setminus I^{\otimes n}$ and a state $|\psi_0\rangle \in \mathcal{S}_n$ such that $\mathcal{M}|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$ for some $\gamma \in \mathbb{R}$, then for any state $|\psi_1\rangle \in \mathcal{R}_{|\psi_0\rangle}$ there exists an $\mathcal{H} \in \mathcal{G}_n \setminus I^{\otimes n}$ such that $\mathcal{H}|\psi_1\rangle = e^{i\gamma}|\psi_1\rangle$.

A consequence of Remark 6 is namely that for any reachable state $|\Psi\rangle$ there are at least two different gates $\mathcal{M}_i, \mathcal{M}_j \in \mathcal{G}_n$ whose corresponding states $\mathcal{M}_i|\psi_0\rangle$ and $\mathcal{M}_j|\psi_0\rangle$ are equivalent (up to a phase) to same state $|\Psi\rangle$, since $\mathcal{M}_i|\psi_0\rangle \equiv \mathcal{M}_j|\psi_0\rangle \equiv |\Psi\rangle$ if we let $|\Psi\rangle = \mathcal{M}_i|\psi_0\rangle$ and $\mathcal{M}_j = \mathcal{H}\mathcal{M}_i$. The number of reachable states $|\mathcal{R}_{|\psi_0\rangle}|$ is thus upper bounded by $1/2|\mathcal{G}_n|$ when starting from a stabilizer state.

B. Continuous, random error accumulation (one-qubit case)

In this section, we analyze the case where a single qubit:

1. receives a random perturbation on the Bloch sphere after each s -th unitary gate according to a continuous distribution $p_s(\alpha)$, and
2. depolarizes to the completely depolarized state $I/2$ with probability $q \in [0, 1]$ after each unitary gate,

by considering it an absorbing random walk on the Bloch sphere. The key point leveraged here is that the trace distance is invariant under rotations. Hence a sufficiently symmetric random walk distribution will give the error probabilities.

a. Model. Let R_0 be an initial point on the Bloch sphere. Every time a unitary quantum gate is applied, the qubit is rotated and receives a small perturbation. This results in a random walk $\{R_t\}_{t \geq 0}$ on the Bloch sphere for as long as the qubit has not depolarized. Because the trace distance is invariant under rotations and since the rotations are applied both to ρ_t and σ_t , we can ignore the rotations. We let ν denote the random time at which the qubit depolarizes. With the usual independence assumptions, $\nu \sim \text{Geometric}(q)$.

Define $\mu_t(r)$ for $t < \nu$ as the probability that the random walk is in a solid angle Ω about r (in spherical coordinates) conditional on the qubit not having depolarized yet. That is,

$$\mathbb{P}[R_t \in \mathcal{S} \mid \nu > t] \triangleq \int_{\mathcal{S}} \mu_t(r) d\Omega(r).$$

We assume without loss of generality that $R_0 = \hat{z}$. From [24], the initial distribution is then given by

$$\mu_0 = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} P_n(\cos \theta).$$

Here, the $P_n(\cdot)$ denote the Legendre polynomials. Also introduce the shorthand notation

$$\Lambda_{n,t} \triangleq \prod_{s=1}^t \int_0^\pi P_n(\cos \alpha) dp_s(\alpha).$$

In particular: if $p_t(\alpha) = \delta(\alpha)$ for all $t \geq 0$, then $\Lambda_{n,t} = (P_n(\cos \alpha))^t$. From [24], it follows that after t unitary quantum gates have been applied without depolarization having occurred,

$$\mu_t = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \Lambda_{n,t} P_n(\cos \theta). \quad (19)$$

b. Results. In this section we specify D_t as the trace distance. We are now in position to state our findings:

Proposition 7 (Single qubit). For $0 \leq \delta \leq 1$, $t \in \mathbb{N}_+$: the expected trace distance satisfies

$$\mathbb{E}[D_t] = \frac{1}{2} - (1-q)^t \left(\frac{1}{2} + 2 \sum_{n=0}^{\infty} \frac{\Lambda_{n,t}}{(2n-1)(2n+3)} \right). \quad (20)$$

The distribution of the trace distance is given by

$$\begin{aligned} \mathbb{P}[D_t \leq \delta] &= \mathbb{1}[\tfrac{1}{2} \in [0, \delta]] (1 - (1-q)^t) + \\ & (1-q)^t \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \sum_{r=1}^{n+1} (-1)^{r+1} \delta^{2r} C_{r-1} \binom{n+r-1}{2(r-1)}. \end{aligned} \quad (21)$$

Here, the C_r denote the Catalan numbers.² Finally; the distribution of maximum trace distance is lower bounded by

$$\mathbb{P}[\max_{0 \leq s \leq t} D_s \leq \delta | \nu > t] \geq 0 \vee \left(1 - t + \delta^2 \sum_{s=1}^t \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,s} \frac{n!}{(2)_n} P_n^{(1,-1)}(1-2\delta^2)\right). \quad (22)$$

Proof of (20). By the law of total expectation, we have

$$\mathbb{E}[D_t] = \mathbb{E}[D_t | \nu > t] \mathbb{P}[\nu > t] + \mathbb{E}[D_t | \nu \leq t] \mathbb{P}[\nu \leq t].$$

Since $\nu \sim \text{Geometric}(q)$, we have that

$$\mathbb{P}[\nu > t] = 1 - \mathbb{P}[\nu \leq t] = (1-q)^t.$$

Note additionally that $D_t = 1/2$ whenever $t \geq \nu$. Therefore

$$\begin{aligned} \mathbb{E}[D_t] &= \mathbb{E}[D_t | \nu > t] (1-q)^t + \frac{1}{2} (1 - (1-q)^t) \\ &= \frac{1}{2} + (\mathbb{E}[D_t | t < \nu] - \frac{1}{2}) (1-q)^t. \end{aligned}$$

We now calculate $\mathbb{E}[D_t | \nu > t]$ using (19) and the Bloch sphere representation:

$$\begin{aligned} \mathbb{E}[D_t | \nu > t] &= \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \Lambda_{n,t} \int_0^\pi 2\pi \sin \theta \sin \frac{\theta}{2} P_n(\cos \theta) d\theta \\ &= \sum_{n=0}^{\infty} \frac{2n+1}{2} \Lambda_{n,t} \int_{-1}^1 \sqrt{\frac{1-x}{2}} P_n(x) dx. \end{aligned} \quad (23)$$

Also recall two facts about the Legendre polynomials: the recurrence relation in [42] states that

$$P_n(x) = \frac{1}{2n+1} (P'_{n+1}(x) - P'_{n-1}(x)), \quad (24)$$

and Rodrigues formula [43, (8.6.18)] states that

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n. \quad (25)$$

Using (24), (25), and integration by parts, we then obtain

$$\begin{aligned} &\int_{-1}^1 \sqrt{\frac{1-x}{2}} P_n(x) dx \\ &= \frac{1}{2n+1} \left(- \int_{-1}^1 \frac{P_{n+1}(x)}{2\sqrt{2-2x}} dx + \int_{-1}^1 \frac{P_{n-1}(x)}{2\sqrt{2-2x}} dx \right). \end{aligned} \quad (26)$$

² Alternative forms include:

$\mathbb{P}[D_t \leq \delta | \nu > t] = \delta^2 \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} {}_2F_1(-n, n+1, 2; \delta^2)$, and $\mathbb{P}[D_t \leq \delta | \nu > t] = \delta^2 \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \frac{n!}{(2)_n} P_n^{(1,-1)}(1-2\delta^2)$ with ${}_2F_1(a, b, c; z)$ the Hypergeometric function, $(\cdot)_n$ the Pochhammer symbol, and $P_n^{(\alpha, \beta)}(x)$ the Jacobi polynomials.

We have by [44, (12.4)] that the generating function of the Legendre polynomials is given by

$$\sum_{m=0}^{\infty} P_m(x) s^m = \frac{1}{\sqrt{1-2xs+s^2}}. \quad (27)$$

Based on (27) with $t = 1$ and the orthogonality of Legendre polynomials,

$$\begin{aligned} \int_{-1}^1 \frac{P_n(x)}{\sqrt{2-2x}} dx &= \int_{-1}^1 P_n(x) \sum_{m=0}^{\infty} P_m(x) dx \\ &= \sum_{m=0}^{\infty} \int_{-1}^1 P_n(x) P_m(x) dx = \frac{2}{2n+1}. \end{aligned} \quad (28)$$

Here, we have used Lebesgue's dominated convergence theorem with $|P_n(x)| \leq 1 \forall n$. Therefore, continuing from (23) using (26) and (28),

$$\begin{aligned} \mathbb{E}[D_t | \nu > t] &= \sum_{n=0}^{\infty} \frac{2n+1}{2} \Lambda_{n,t} \left(- \int_{-1}^1 \frac{P_{n+1}(x)}{2\sqrt{2-2x}} dx \right. \\ &\quad \left. + \int_{-1}^1 \frac{P_{n-1}(x)}{2\sqrt{2-2x}} dx \right) \\ &= \sum_{n=0}^{\infty} \frac{2n+1}{2} \Lambda_{n,t} \frac{-4}{(2n-1)(2n+1)(2n+3)}. \end{aligned}$$

Simplifying gives the result.

Proof of (21). Similar to above we have by the law of total probability that

$$\begin{aligned} \mathbb{P}[a \leq D_t \leq b] &= \mathbb{P}[a \leq D_t \leq b | \nu \leq t] \mathbb{P}[\nu \leq t] \\ &\quad + \mathbb{P}[a \leq D_t \leq b | \nu > t] \mathbb{P}[\nu > t], \end{aligned}$$

and we note now that $\mathbb{P}[a \leq D_t \leq b | \nu \leq t] = \mathbb{1}[\frac{1}{2} \in [a, b]]$. Therefore

$$\begin{aligned} \mathbb{P}[a \leq D_t \leq b] &= \mathbb{1}[\frac{1}{2} \in [a, b]] (1 - (1-q)^t) \\ &\quad + \mathbb{P}[a \leq D_t \leq b | \nu > t] (1-q)^t. \end{aligned}$$

We now calculate $\mathbb{P}[a \leq D_t \leq b | \nu > t]$; again using (19). Let $0 \leq a \leq b \leq 1$. From the equivalence of the events

$$\{a \leq D_t \leq b\} = \{2 \arcsin(a) \leq \Theta_t \leq 2 \arcsin(b)\},$$

where Θ_t denotes the polar angle of R_t , it follows that

$$\begin{aligned} \mathbb{P}[a \leq D_t \leq b] &= (1 - (1-q)^t) \mathbb{1}[\frac{1}{2} \in [a, b]] \\ &\quad + (1-q)^t \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \Lambda_{n,t} \int_{2 \arcsin a}^{2 \arcsin b} 2\pi \sin \theta P_n(\cos \theta) d\theta. \end{aligned} \quad (29)$$

Now let $0 \leq \delta \leq 1$. Continuing from (29), since $\cos(2 \arcsin \delta) = 1 - 2\delta^2$ for $\delta \in [0, 1]$ and letting $\cos \theta = x$,

$$\mathbb{P}[D_t \leq \delta | \nu > t]$$

$$\begin{aligned}
&= \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \Lambda_{n,t} \int_0^{2\arcsin \delta} 2\pi \sin \theta P_n(\cos \theta) d\theta \\
&= \sum_{n=0}^{\infty} \frac{2n+1}{2} \Lambda_{n,t} \int_{1-2\delta^2}^1 P_n(x) dx.
\end{aligned}$$

By the explicit representation of Rodrigues' formula [43, (8.6.18)],

$$\begin{aligned}
&\mathbb{P}[D_t \leq \delta | \nu > t] \\
&= \sum_{n=0}^{\infty} \frac{2n+1}{2} \Lambda_{n,t} \int_{1-2\delta^2}^1 \sum_{k=0}^n \binom{n}{k} \binom{n+k}{k} \left(\frac{x-1}{2}\right)^k dx \\
&= \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \sum_{k=0}^n \binom{n}{k} \binom{n+k}{k} \frac{(-1)^k}{k+1} \delta^{2(k+1)}.
\end{aligned}$$

Finally, let $r = k + 1$, such that

$$\begin{aligned}
&\mathbb{P}[D_t \leq \delta | \nu > t] \\
&= \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \sum_{r=1}^{n+1} \binom{n}{r-1} \binom{n+r-1}{r-1} \frac{(-1)^{r-1}}{r} \delta^{2r} \\
&= \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \sum_{r=1}^{n+1} (-1)^{r-1} \delta^{2r} C_{r-1} \binom{n+r-1}{2(r-1)}.
\end{aligned}$$

Proof of (22). This follows directly after applying De Morgan's law and Boole's inequality, i.e.,

$$\begin{aligned}
&\mathbb{P}[\max_{0 \leq s \leq t} D_s \leq \delta | \nu > t] = \mathbb{P}\left[\bigcap_{s=0}^t \{D_s \leq \delta\} \middle| \nu > t\right] \\
&= \mathbb{P}\left[\left(\bigcup_{s=0}^t \{D_s > \delta\}\right)^c \middle| \nu > t\right] = 1 - \mathbb{P}\left[\bigcup_{s=0}^t \{D_s > \delta\} \middle| \nu > t\right] \\
&\geq 1 - \sum_{s=0}^t \mathbb{P}[D_s > \delta | \nu > t] = 1 - t + \sum_{s=0}^t \mathbb{P}[D_s \leq \delta | \nu > t].
\end{aligned}$$

That is it. \square

IV. SIMULATIONS

A. Error accumulation in randomized benchmarking

We now consider error accumulation in single-qubit randomized benchmarking. In each randomized benchmarking simulation experiment, the initial state is set to $|1\rangle$ and subsequently $\tau - 1$ gates are selected one by one from the Clifford group \mathcal{C}_1 uniformly at random. Finally, based on the experimental setup in [12], we add a τ -th gate that transfers the state to $|0\rangle$ in the absence of errors. For simplicity we specify $p = 1$ and thus discuss the trace distance throughout this section.

1. Pauli and Clifford channel errors

We consider two kinds of error models: Pauli channels and Clifford channels. For the Pauli channel model, let the probability that no noise occurs be $\mathbb{P}(\Lambda = I) = 1 - r$, and the probabilities of every noise type occurring be $\mathbb{P}(\Lambda = X) = \mathbb{P}(\Lambda = Y) = \mathbb{P}(\Lambda = Z) = r/3$, where $r \in [0, 1]$. For the Clifford channel model, let the probability of no noise occurring be $\mathbb{P}(\Lambda = I) = 1 - r$, and the probabilities of every other gate type occurring equal $r/23$. In Figure 4 the parameter r is set to $1/100$. Two error thresholds δ are considered: $\delta = 1/10$ (a, c, and d) and $\delta = 1/5$ (b). The insets show the influence of parameter r on the probability of error in (6) and the probability of maximum error in (7) at time $t = 100$. The results in Figure 4 illustrate the theoretical results for the probability of error (6), the expectation of the trace distance, and the probability of maximum error (7), and their validity is supported by these simulations. Figure 4 also illustrates that different error models lead to different error accumulation behaviors.

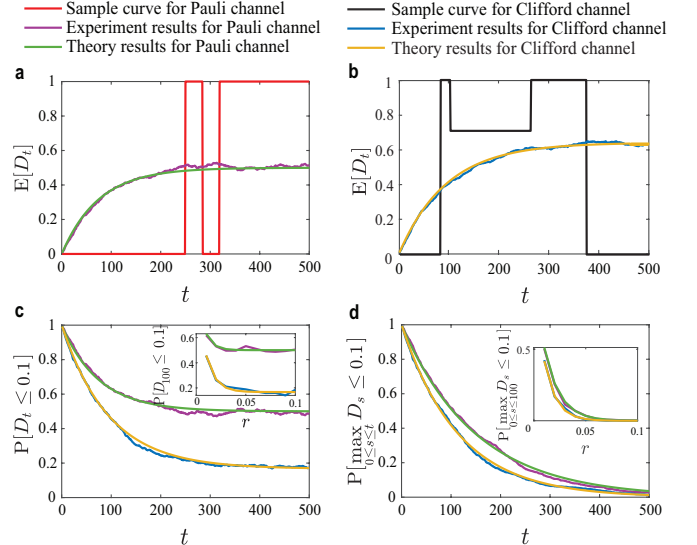


FIG. 4. The error accumulation based on Pauli and Clifford channels in randomized benchmarking. Two error thresholds δ are considered, $\delta = 1/10$ (figures a, c, d) and $\delta = 1/5$ (figure b). The simulation results are calculated from 1000 independent randomized benchmarking experiments.

2. Influence of the initial state on Pauli error accumulation

In this section we consider the influence of the initial state on Pauli error accumulation. We ignore the last gate of randomized benchmarking for simplicity. Each gate is selected one by one from the Pauli group uniformly at random. The error model described above is considered again and the parameter r is set to $1/5$.

Figure 5 shows the state transition diagram for two different initial states: $|\zeta_0\rangle = \sqrt{7/10}|0\rangle + \sqrt{3/10}|1\rangle$ and $|\xi_0\rangle = \sqrt{4/5}|0\rangle + \sqrt{1/5}|1\rangle$. The bad state pairs that constitute $\mathcal{B}_{|\zeta_0\rangle, \delta}^{|\zeta_0\rangle}$ and $\mathcal{B}_{|\xi_0\rangle, \delta}^{|\xi_0\rangle}$, which have a trace distance over $\delta = 1/5$, are indicated in red. Note that the number of bad state pairs can be affected by the choice of initial state. Figure 6 shows the probability of maximum error in (7) and the maximum number of tolerant gates in (8) for the same two different initial states: $|\zeta_0\rangle$ (upper) and $|\xi_0\rangle$ (bottom). Figure 5 and Figure 6 illustrate too that the choice of initial state can affect the probability $\mathbb{P}[\max_{0 \leq s \leq t} D_s > \delta]$ and the maximum number of tolerant gates $t_{\delta, \gamma}^*$. Finally, when starting from the initial state $|\zeta_0\rangle$, in this simple case, (7) reduces to

$$\mathbb{P}[\max_{0 \leq s \leq t} D_s > 1/5] = 1 - (1 - \frac{2}{3}r)^t,$$

while when starting from the initial state $|\xi_0\rangle$ we have

$$\mathbb{P}[\max_{0 \leq s \leq t} D_s > 1/5] = 1 - (1 - r)^t.$$

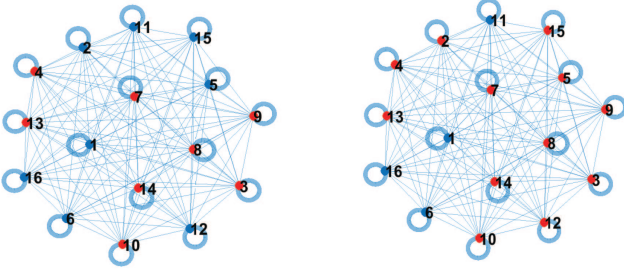


FIG. 5. State transition diagram for different initial states: $|\zeta_0\rangle$ (left) and $|\xi_0\rangle$ (right), and the error threshold $\delta = 1/5$. The red nodes show the bad state pairs in $\mathcal{B}_{|\zeta_0\rangle, \delta}^{|\zeta_0\rangle}$ and $\mathcal{B}_{|\xi_0\rangle, \delta}^{|\xi_0\rangle}$, respectively, in which the trace distances are larger than δ .

B. Error accumulation in nonrandom circuits

Here we illustrate error accumulation rates in two non-random circuits. One is a periodical single-qubit circuit that repeats a Hadamard, Pauli- X , Pauli- Y and Pauli- Z gate $k = 25$ times. Another is the periodical two-qubit circuit shown in Figure 7 repeated $k = 5$ times. Here the controlled-NOT gate

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

We also consider here an error model in which the errors are independent from the gates. The error model in the single-qubit circuit is as follows:

$$\mathbb{P}(\Lambda = I) = 0.990, \quad \mathbb{P}(\Lambda = Z) = 0.010.$$

The error model in the two-qubit circuit, when labeling the qubits by A and B , is chosen as:

$$\mathbb{P}(\Lambda_A = I) = 0.990, \quad \mathbb{P}(\Lambda_A = X) = 0.006,$$

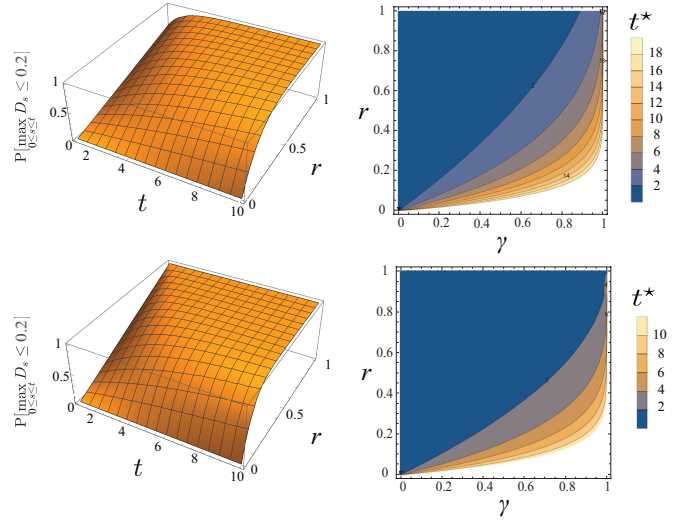


FIG. 6. Pauli channel error accumulation on single-qubit randomized benchmarking when starting from different initial states: $|\zeta_0\rangle$ (top) and $|\xi_0\rangle$ (bottom). The error threshold is set to $\delta = 1/5$.

$$\begin{aligned} \mathbb{P}(\Lambda_A = Y) &= 0.003, \quad \mathbb{P}(\Lambda_A = Z) = 0.001, \\ \mathbb{P}(\Lambda_B = I) &= 0.980, \quad \mathbb{P}(\Lambda_B = X) = 0.002, \\ \mathbb{P}(\Lambda_B = Y) &= 0.014, \quad \mathbb{P}(\Lambda_B = Z) = 0.004. \end{aligned}$$

The error threshold is set to $\delta = 1/10$. The theoretical and simulation results on the two circuits are shown in Figure 7. Note that the simulation curves almost coincide with the theoretical curves; the deviation is only due to numerical limits. From a computational point of view, the theoretical results take only about 1% of the running time of the simulation results which are calculated from 2000 independent repetitions. This shows that our theoretical method can actually give more accurate results with less computing effort. However, we expect available memory to be the primary constraining factor when numerically implementing our formulae. Note also that because different gates influence error accumulation to different degrees, the periodical ladder-like behavior occurs in Figure 7.

C. Continuous, random error accumulation in a single qubit

We now simulate the accumulation of continuous errors without depolarization ($q = 0$) in a single qubit. Here, the noise is assumed to lead to a random walk on the Bloch sphere that takes steps of a fixed angle $\alpha = 1/10$, and therefore $P_t(\alpha) = \delta(\alpha)$. The threshold δ is set to be $1/10$. The theoretical mean trace distance $\mathbb{E}[D_t]$ and probability $\mathbb{P}[D_t \leq \delta]$ are calculated using (20) and (21). The theoretical results and simulations are shown in Figure 8.

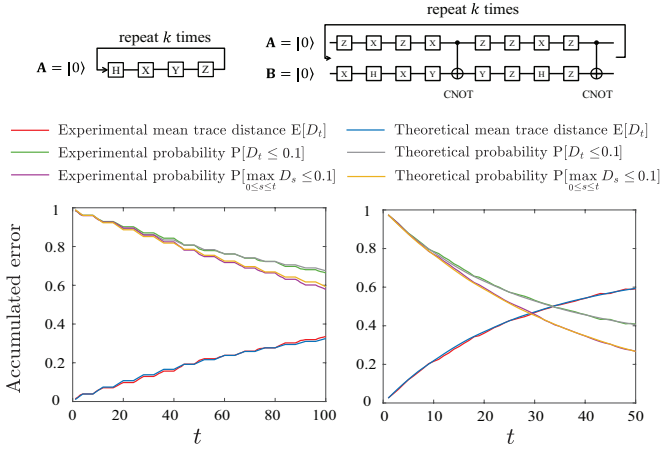


FIG. 7. Theoretical and simulation results for error accumulation on a single-qubit circuit (left) and a two-qubit circuit (right). The numerical results are calculated from 2000 independent runs, and almost indistinguishable from the formulae.

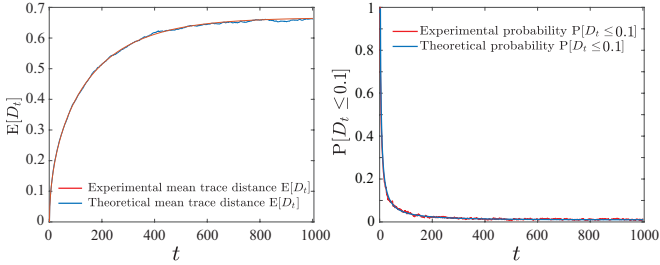


FIG. 8. Continuous error accumulation in one qubit. The numerical results are from 2000 independent runs of our simulation.

V. FAULT-TOLERANT QUANTUM CIRCUIT OPTIMIZATION

The rate at which errors accumulate may be different for different quantum circuits that can implement the same algorithm. Using techniques from optimization and (14), we can therefore search for the quantum circuit that has the lowest error rate accumulation while maintaining the same final state. To see this, suppose we are given a circuit $\mathcal{U}_\tau = \{U_1, U_2, \dots, U_\tau\}$. For given ρ_0 this brings the quantum state to some quantum state ρ_τ . Other circuits may go to the same final state and have a lower probability of error at time τ . We will therefore aim to

$$\begin{aligned} & \underset{G_1, \dots, G_\tau \in \mathcal{G}_n}{\text{minimize}} && u(\{G_1, \dots, G_\tau\}) \\ & \text{subject to} && G_\tau \cdots G_1 = U_\tau \cdots U_1. \end{aligned} \quad (30)$$

Here, one can for example choose for the objective function $u(\cdot)$ the probability of error (13), or probability of maximum error (14). To solve (30), we design a simulated annealing algorithm to improve the quantum circuit.

A. Simulated annealing

We will generate candidate circuits as follows. Let

$$\{G_1^{[\eta]}, \dots, G_\tau^{[\eta]}\}$$

denote the circuit at iteration η . Choose an index $I \in [\tau - 1]$ uniformly at random, choose $G \in \mathcal{G}$ uniformly at random. Then set

$$G_i^{[\eta+1]} = \begin{cases} G & \text{if } i = I, \\ G \leftarrow G_I^{[\eta]} G_{I+1}^{[\eta]} & \text{if } i = I + 1, \\ G_i^{[\eta]} & \text{otherwise.} \end{cases}$$

Let

$$E = \{ \{G_1, \dots, G_\tau\} | G_\tau \cdots G_1 = U_\tau \cdots U_1 \}$$

denote the set of all viable circuits. For two arbitrary circuits $i, j \in E$, let

$$\Delta(i, j) \triangleq \sum_{s=1}^{\tau-1} \mathbb{1}[i_s \neq j_s, i_{s+1} \neq j_{s+1}]$$

denote the number of consecutive gates that differ between both circuits. Under this construction, the *candidate-generator matrix* is given by

$$q_{ij} = \begin{cases} \frac{1}{(\tau-1)|\mathcal{G}|} & \text{if } \Delta(i, j) \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

We will use the Metropolis algorithm. Since the candidate-generator matrix is symmetric, this algorithm means that we set

$$\alpha_{i,j}(T) = \exp\left(-\frac{1}{T} \max\{0, u(j) - u(i)\}\right)$$

as the *acceptance probability* of circuit j over i . Here $T \in (0, \infty)$ is a positive constant. Finally, we need a cooling schedule. Let

$$M \triangleq \sup_{\{i, j \in E | \Delta(i, j) \leq 1\}} \{U(j) - U(i)\}.$$

Based on [23], if we choose a cooling schedule $\{T_\eta\}_{\eta \geq 0}$ that satisfies

$$T_\eta \geq \frac{\tau M}{\ln \eta},$$

then the Metropolis algorithm will converge to the set of global minima of the minimization problem in (30).

Lemma 8. *Algorithm 1 converges to the global minimizer of (30) whenever $T_\eta \geq \tau M / \ln \eta$ for $\eta = 1, 2, \dots$*

Input: A group \mathcal{G} , a circuit $\{U_1, \dots, U_\tau\}$, and number of iterations w
Output: A revised circuit $\{G_1^{[w]}, \dots, G_\tau^{[w]}\}$
begin
 Initialize $\{G_1^{[0]}, \dots, G_\tau^{[0]}\} = \{U_1, \dots, U_\tau\}$;
 for $\eta \leftarrow 1$ **to** w **do**
 Choose $I \in [\tau - 1]$ uniformly at random;
 Choose $G \in \mathcal{G}$ uniformly at random;
 Set $J_I = G, J_{I+1} = G^\dagger G_I^{[\eta]} G_{I+1}^{[\eta]}, J_i = G_i^{[\eta]} \forall i \neq I, I+1$;
 Choose $X \in [0, 1]$ uniformly at random;
 if $X \leq \alpha_{G^{[\eta]}, J}^{(T_\eta)}$ **then**
 Set $G^{[\eta+1]} = J$;
 else
 Set $G^{[\eta+1]} = G^{[\eta]}$;
 end
 end
end

Algorithm 1: Pseudo-code for the simulated annealing algorithm described in Section V–A.

B. Examples

1. Gate-dependent error model

We are going to improve the one-qubit circuit in Figure 7 using Algorithm 1. The gates are limited to the Clifford group \mathcal{C}_1 and the errors will be limited to the Pauli channel. The error probabilities considered here are gate-dependent and can be found in §E. The cooling schedule used here will be set as $T_\eta = C/\ln(\eta + 1)$, and the algorithm's result when using $C = 0.004$ is shown in Figure 9. Figure 9 illustrates that the improved circuit can indeed lower the error accumulation rate. The circuit with the lowest error accumulation rate that was found is shown in §H.

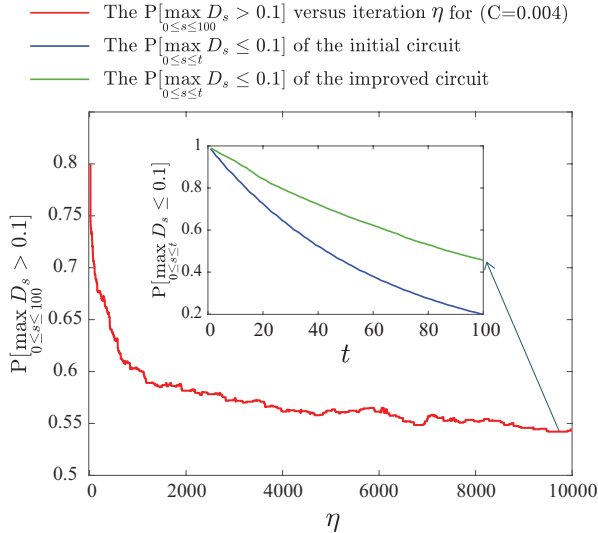


FIG. 9. Circuit optimization when using Algorithm 1. The error probabilities are gate-dependent. Note that the probability of maximum error (14) decreases as the number of iterations η increases when using Algorithm 1 ($C = 0.004$). Here we started from the one-qubit circuit in Figure 7.

2. Gates in a subset of one group

The gates that are available in practice may be restricted to some subset $\mathcal{A} \subseteq \mathcal{G}$. Under such constraint, we could generate candidate circuits as follows: Let $\{G_1^{[\eta]}, \dots, G_\tau^{[\eta]}\}$ denote the circuit at iteration η . In each iteration, two neighboring gates will be considered to be replaced by two other neighboring gates. There are $m \leq (\tau - 1)$ neighboring gate pairs $(G_1^{[\eta]}, G_2^{[\eta]}), \dots, (G_{m-1}^{[\eta]}, G_m^{[\eta]})$ that can be replaced by two different neighboring gates. Choose an index $I \in [m - 1]$ uniformly at random, and replace $(G_I^{[\eta]}, G_{I+1}^{[\eta]})$ by any gate pair from $\{(\tilde{G}_1, \tilde{G}_2) \in \mathcal{A}^2 \mid G_I^{[\eta]} G_{I+1}^{[\eta]} = \tilde{G}_1 \tilde{G}_2\}$ uniformly at random. Pseudo-code for this modified algorithm can be found in §G. It must be noted that this algorithm is not guaranteed to converge to the global minimizer of (30) (due to limiting the gates available); however, it may still find use in practical scenarios where one only has access to a restricted set of gates.

We now aim to decrease the probability of maximum error (14) by changing the two-qubit circuit shown in Figure 7. The error model is the same as that in Section IV–B. The set of gates available for improving the circuit is here limited to $\{I, X, Y, Z, H, CNOT\}$. The result here for the two-qubit circuit is obtained by again using the cooling schedule $T_\eta = C/\ln(\eta + 1)$ but now letting the parameter $C = 0.002$. Figure 10 shows that a more error-tolerant circuit can indeed be found using this simulated annealing algorithm. The improved circuit is shown in §H.

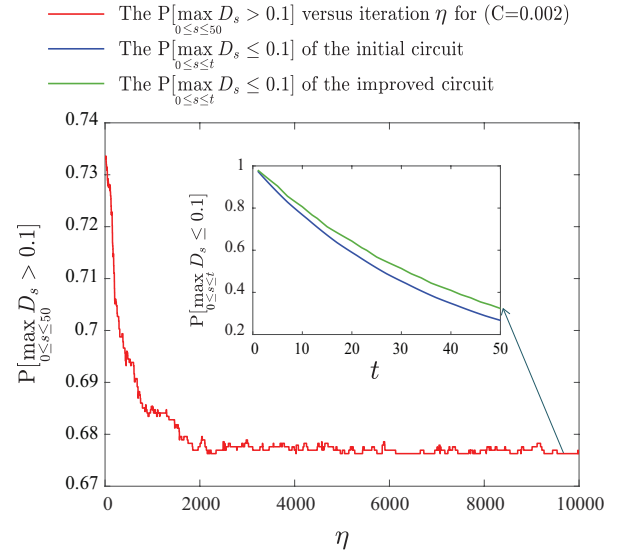


FIG. 10. Circuit optimization when using Algorithm 2. The set of gates available is chosen limited to $\{I, X, Y, Z, H, CNOT\}$. Note that the probability of maximum error (14) decreases as the number of iterations η increases when using Algorithm 2 ($C = 0.002$). Here we started from the two-qubit circuit shown in Figure 7.

VI. CONCLUSION

In conclusion; we have proposed and studied a model for discrete Markovian error accumulation in a multi-qubit quantum computation, as well as a model describing continuous errors accumulating in a single qubit. By modeling the quantum computation with and without errors as two coupled Markov chains, we were able to capture a weak form of time-dependency, allow for fairly generic error distributions, and describe multi-qubit systems. Furthermore, by using techniques from discrete probability theory, we could calculate the probability that error measures such as the fidelity and trace distance exceed a threshold analytically. To combat the numerical challenge that may occur when evaluating our expressions, we additionally provided an analytical bound on the error probabilities that is of lower numerical complexity, and we also discussed the state space reduction that occurs for stabilizer circuits. Finally, we showed how our expressions can be used to decide how many gates one can apply before too many errors accumulate with high probability, and how one can lower the rate of error accumulation in existing circuits by using techniques from optimization.

The bridging of techniques from probability theory and operations research to the domain of quantum computing is novel and presents a new area of research. This paper lay down a foundation for one error accumulation model, and multiple interesting follow-up topics can now be investigated as future research. Here, we provide four intriguing ideas:

- The accumulation of errors when using a universal gate set would need to be modeled using stochastic processes that live on infinite state spaces. Such an approach looks to be connected to the modeling of random walks on

manifolds. This would be a challenging, intriguing, and important next step for the analysis of error accumulation in quantum circuits.

- The expressions in (7) and (14) are, essentially, generalized forms of a geometric distribution. For particular groups and error models, it may be that this expression is well-approximated by a standard geometric distribution (which would be of substantially lower numerical complexity). It would be interesting to investigate whether a reduction of (7) and (14) occurs, or whether an approximation can be found, for particular quantum systems.

- With that idea in mind, note that the hitting time of the set $\mathcal{B}_{|\psi_0\rangle, \delta}^{|\Psi_0\rangle}$ is naturally related to its size relative to the size of the group \mathcal{G}_n . As the number of qubits increases, both of these sets grow in size. Investigating the growth relation between these two sets for particular groups via e.g. techniques from analytical combinatorics [45] may reveal an asymptotic distributional law for the errors in quantum computations with many qubits.

- The availability of an analytical expression for the accumulation of errors allows us to proceed with second-tier optimization methods. For example, any quantum computer architecture would, to achieve practical quantum computing in the near future, have some classical control mechanism that routinely takes operational decisions: which gate do we apply next, do we now apply an error correction procedure, etc. Each of these different operations has its own cost associated with it, e.g. in the form of classical compute time or the loss of ancillary qubits. Using techniques from decision theory [46], we can weigh the long-term effects of different operations through the available analytical expressions, and we could overall achieve more efficient computations in the future. Essentially, we could then compute more with fewer qubits.

-
- [1] Michael A. Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.
 - [2] John Preskill. Quantum computing: pro and con. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):469–486, 1998.
 - [3] Daniel Gottesman. Efficient fault tolerance. *Nature*, 540:44, 2016.
 - [4] Julia Cramer, Norbert Kalb, M. Adriaan Rol, Bas Hensen, Machiel S. Blok, Matthew Markham, Daniel J. Twitchen, Ronald Hanson, and Tim H. Taminiau. Repeated quantum error correction on a continuously encoded qubit by real-time feedback. *Nature communications*, 7:11526, 2016.
 - [5] Norbert M. Linke, Mauricio Gutierrez, Kevin A. Landsman, Caroline Figgatt, Shantanu Debnath, Kenneth R. Brown, and Christopher Monroe. Fault-tolerant quantum error detection. *Science advances*, 3(10):e1701074, 2017.
 - [6] Austin G. Fowler and Lloyd C.L. Hollenberg. Scalability of Shors algorithm with a limited set of rotation gates. *Physical Review A*, 70(3):032329, 2004.
 - [7] Daniel Greenbaum and Zachary Dutton. Modeling coherent errors in quantum error correction. *Quantum Science and Technology*, 3(1):015007, 2017.
 - [8] Easwar Magesan, Daniel Puzzuoli, Christopher E. Granade, and David G. Cory. Modeling quantum noise for efficient testing of fault-tolerant circuits. *Physical Review A*, 87(1):012324, 2013.
 - [9] Mauricio Gutiérrez, Lukas Svec, Alexander Vargo, and Kenneth R. Brown. Approximation of realistic errors by Clifford channels and Pauli measurements. *Physical Review A*, 87(3):030302, 2013.
 - [10] Sergey Bravyi, Matthias Englbrecht, Robert König, and Nolan Peard. Correcting coherent errors with surface codes. *npj Quantum Information*, 4(1):55, 2018.
 - [11] Eric Huang, Andrew C. Doherty, and Steven Flammia. Performance of quantum error correction with coherent errors. *Physical Review A*, 99(2):022313, 2019.
 - [12] T. Xia, M. Lichtman, K. Maller, A.W. Carr, M.J. Piotrowicz, L. Isenhowe, and M. Saffman. Randomized benchmarking of single-qubit gates in a 2D array of neutral-atom qubits. *Physical Review Letters*, 114(10):100503, 2015.
 - [13] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quan-

- tum processes. *Physical Review Letters*, 106(18):180504, 2011.
- [14] Harrison Ball, Thomas M. Stace, Steven T. Flammia, and Michael J. Biercuk. Effect of noise correlations on randomized benchmarking. *Physical Review A*, 93(2):022303, 2016.
 - [15] Smitha Janardan, Yu Tomita, Mauricio Gutiérrez, and Kenneth R. Brown. Analytical error analysis of Clifford gates by the fault-path tracer method. *Quantum Information Processing*, 15(8):3065–3079, 2016.
 - [16] Joel J. Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, January 2018.
 - [17] Jeffrey M. Epstein, Andrew W. Cross, Easwar Magesan, and Jay M. Gambetta. Investigating the limits of randomized benchmarking protocols. *Physical Review A*, 89(6):062321, 2014.
 - [18] Daniel Stilck França and A.K. Hashagen. Approximate randomized benchmarking for finite groups. *Journal of Physics A: Mathematical and Theoretical*, 51(39):395302, 2018.
 - [19] Winton G. Brown and Bryan Eastin. Randomized benchmarking with restricted gate sets. *Physical Review A*, 97(6):062323, 2018.
 - [20] Bryan H. Fong and Seth T. Merkel. Randomized benchmarking, correlated noise, and ising models. *arXiv preprint arXiv:1703.09747*, 2017.
 - [21] Christopher J. Wood and Jay M. Gambetta. Quantification and characterization of leakage errors. *Physical Review A*, 97(3):032306, 2018.
 - [22] Joel J. Wallman, Marie Barnhill, and Joseph Emerson. Robust characterization of loss rates. *Physical Review Letters*, 115(6):060501, 2015.
 - [23] Pierre Brémaud. *Discrete probability models and methods*, volume 78. Springer, 2017.
 - [24] Paul Harry Roberts and Harold Douglas Ursell. Random walk on a sphere and on a Riemannian manifold. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 252(1012):317–356, 1960.
 - [25] Henryk Gutmann. *Description and control of decoherence in quantum bit systems*. PhD thesis, lmu, 2005.
 - [26] Emanuel Knill. Quantum computing with realistically noisy devices. *Nature*, 434(7029):39, 2005.
 - [27] Robin Harper, Ian Hincks, Chris Ferrie, Steven T. Flammia, and Joel J. Wallman. Statistical analysis of randomized benchmarking. *Physical Review A*, 99(5):052350, 2019.
 - [28] Nikolaž Moll, Panagiotis Barkoutsos, Lev S. Bishop, Jerry M. Chow, Andrew Cross, Daniel J. Egger, Stefan Filipp, Andreas Fuhrer, Jay M. Gambetta, Marc Ganzhorn, et al. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology*, 3(3):030503, 2018.
 - [29] Keisuke Fujii. Stabilizer formalism and its applications. In *Quantum Computation with Topological Codes*, pages 24–55. Springer, 2015.
 - [30] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004.
 - [31] Daniel M. Greenberger, Michael A. Horne, and Anton Zeilinger. Going beyond Bells theorem. In *Bells theorem, quantum theory and conceptions of the universe*, pages 69–72. Springer, 1989.
 - [32] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states. *Physical Review Letters*, 69(20):2881, 1992.
 - [33] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Physical Review Letters*, 70(13):1895, 1993.
 - [34] Peter Selinger. Generators and relations for n-qubit Clifford operators. *Logical Methods in Computer Science*, 11, 2013.
 - [35] Daniel Gottesman. The Heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.
 - [36] Mary Beth Ruskai. Pauli exchange errors in quantum computation. *Physical Review Letters*, 85(1):194, 2000.
 - [37] Robert Koenig and John A. Smolin. How to efficiently select an arbitrary Clifford group element. *Journal of Mathematical Physics*, 55(12):122202, 2014.
 - [38] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.
 - [39] Piet Van Mieghem. *Performance analysis of complex networks and systems*. Cambridge University Press, 2014.
 - [40] Daniel Gottesman. Stabilizer codes and quantum error correction. *arXiv preprint quant-ph/9705052*, 1997.
 - [41] Héctor J. García, Igor L. Markov, and Andrew W. Cross. On the geometry of stabilizer states. *arXiv preprint arXiv:1711.07848*, 2017.
 - [42] C.C. Grosjean. Theory of recursive generation of systems of orthogonal polynomials: An illustrative example. *Journal of Computational and Applied Mathematics*, 12:299–318, 1985.
 - [43] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1965.
 - [44] George B. Arfken and Hans J. Weber. *Mathematical methods for physicists*, 1999.
 - [45] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. cambridge University press, 2009.
 - [46] Martin L. Puterman. *Markov decision processes: Discrete stochastic dynamic programming*. 1994.
 - [47] Steven Roman, S. Axler, and F.W. Gehring. *Advanced linear algebra*, volume 3. Springer, 2005.

Acknowledgments. This research received financial support from the Chinese Scholarship Council (CSC) in the form of a CSC Scholarship.

Appendix A: \mathcal{C}_n is a group

The fact that \mathcal{C}_n is a group can be verified by checking the two necessary properties: (i) $I^{\otimes n} \in \mathcal{C}_n$ because it is unitary and $I^{\otimes n} \sigma (I^{\otimes n})^\dagger = \sigma$. (ii) Suppose $C \in \mathcal{C}_n$, such that for any $\sigma \in \pm P_n^*$, we have that $C \sigma C^\dagger \in \pm P_n^*$. This implies that for any $\omega \in \pm P_n^*$, we can find a $\sigma \in \pm P_n^*$ such that $\omega = C \sigma C^\dagger$. Conclude that because C is unitary,

$C^{-1}\omega(C^{-1})^\dagger = C^\dagger\omega C = C^\dagger C\sigma C C^\dagger = \sigma \in P_n^*$. Hence $C^{-1} \in \mathcal{C}_n$.

Appendix B: Relation between the error probabilities when using the trace distance and fidelity

Let t be s.t. $0 \leq t \leq \tau$ and let $\omega \in \{D_t \leq \varepsilon\} = \{1 - D_t \geq 1 - \varepsilon\}$. By [1, (9.110)], we have that $1 - F_t \leq D_t \leq \sqrt{1 - F_t^2}$ for all $t \geq 0$. Consequentially $1 - D_t \leq F_t \leq \sqrt{1 - D_t^2}$ for all $t \geq 0$. On every such ω , we thus also have that $F_t \geq 1 - \varepsilon$. We have shown that $\{D_t \leq \varepsilon\} \subseteq \{F_t \geq 1 - \varepsilon\}$, which proves the first statement. For the second statement, we similarly note that $\{\min_{0 \leq s \leq t} F_s \geq 1 - \varepsilon\} \supseteq \{\min_{0 \leq s \leq t} (1 - D_s) \geq 1 - \varepsilon\} = \{\max_{0 \leq s \leq t} D_s \leq \varepsilon\}$. \square

Appendix C: Number of stabilizer states for a gate

For n qubits, any gate $\mathcal{M} \in \mathcal{G}_n \setminus I^{\otimes n}$ can be represented using a $2^n \times 2^n$ unitary matrix. Recall that any unitary matrix of finite size is unitarily diagonalizable since every unitary matrix is normal [47]. A $2^n \times 2^n$ matrix that is diagonalizable must have a set of 2^n linearly independent eigenvectors [47].

The initial states $|\psi_0\rangle$ that can satisfy $\mathcal{M}|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$ are the eigenvectors of the matrix \mathcal{M} with eigenvalue $\lambda = e^{i\gamma}$. For any unitary matrix A with eigenvalue λ and eigenvector v , $A^\dagger A = AA^\dagger = I$, $v^\dagger v = v^\dagger A^\dagger A v = \lambda^\dagger v^\dagger v \lambda = \lambda^\dagger \lambda v^\dagger v$. Also recall that any eigenvector $\|v\| \neq 0$ by definition [47] and thus it always holds that $|\lambda| = 1$. So $\mathcal{M}|\psi_0\rangle = \lambda|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$. \square

Appendix D: A stabilizer state follows after a stabilizer state

By assumption and the definition in (18), for any state $|\psi_1\rangle \in \mathcal{R}_{|\psi_0\rangle}$, $\exists \mathcal{Z} \in \mathcal{G}_n : |\psi_1\rangle = \mathcal{Z}|\psi_0\rangle$ since \mathcal{G}_n is a group. we have furthermore that $\exists \mathcal{H} \in \mathcal{G}_n \setminus I^{\otimes n} : \mathcal{H}\mathcal{Z} = \mathcal{Z}\mathcal{M}$. Then $|\psi_1\rangle = \mathcal{Z}|\psi_0\rangle = e^{-i\gamma}\mathcal{Z}\mathcal{M}|\psi_0\rangle = e^{-i\gamma}\mathcal{H}\mathcal{Z}|\psi_0\rangle = e^{-i\gamma}\mathcal{H}|\psi_1\rangle$. So $\mathcal{H}|\psi_1\rangle = e^{i\gamma}|\psi_1\rangle$. \square

Appendix E: Gate-dependent error model

In Table I, we provide the precise error probabilities used in Section IV.

Appendix F: Method to find all reachable stabilizer states

All reachable stabilizer states can be found given the finite unitary group \mathcal{G}_n of gates (and noise) and the initial stabilizer state $|\psi_0\rangle$. Given an initial stabilizer state $|\psi_0\rangle$, the reduced states can be found by the following steps. First list all gates (and noise) $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n\}$ in group \mathcal{G}_n . All reachable states are then $\{\mathcal{M}_1|\psi_0\rangle, \mathcal{M}_2|\psi_0\rangle, \dots, \mathcal{M}_n|\psi_0\rangle\}$. At last, any two states $\mathcal{M}_i|\psi_0\rangle$ and $\mathcal{M}_j|\psi_0\rangle$ that satisfies $\mathcal{M}_i|\psi_0\rangle = e^{i\gamma}\mathcal{M}_j|\psi_0\rangle$ will fall into the same state.

Appendix G: Pseudo-code for gate-limited simulated annealing

In Algorithm 2, we present the pseudo-code for the simulated annealing algorithm when restricting to a subset of available gates.

Appendix H: Improved circuits

In Figure 11, we present the circuits with the lowest error accumulation rates found by our implementations of the two simulated annealing algorithms.

	$\mathbb{P}[\Lambda = I \mid c_i]$	$\mathbb{P}[\Lambda = X \mid c_i]$	$\mathbb{P}[\Lambda = Y \mid c_i]$	$\mathbb{P}[\Lambda = Z \mid c_i]$
c_1	0.990	0.003	0.003	0.003
c_2	0.965	0.0123	0.0103	0.0123
c_3	0.983	0.0043	0.0083	0.0043
c_4	0.977	0.0083	0.0103	0.0043
c_5	0.969	0.0113	0.0073	0.0123
c_6	0.984	0.0063	0.0043	0.0053
c_7	0.979	0.0043	0.013	0.003
c_8	0.987	0.0043	0.0033	0.0053
c_9	0.979	0.003	0.0093	0.0083
c_{10}	0.985	0.0053	0.0053	0.0043
c_{11}	0.980	0.0073	0.003	0.0093
c_{12}	0.975	0.0083	0.0063	0.0103
c_{13}	0.974	0.0113	0.0063	0.0083
c_{14}	0.975	0.0073	0.0063	0.0113
c_{15}	0.972	0.013	0.0093	0.0053
c_{16}	0.980	0.0043	0.0093	0.0063
c_{17}	0.979	0.0063	0.0093	0.0053
c_{18}	0.982	0.0103	0.0043	0.003
c_{19}	0.977	0.0063	0.0043	0.0123
c_{20}	0.976	0.0113	0.0073	0.0103
c_{21}	0.975	0.0073	0.073	0.0103
c_{22}	0.967	0.0073	0.0073	0.0103
c_{23}	0.974	0.013	0.0063	0.0063
c_{24}	0.978	0.0123	0.0053	0.0043

TABLE I. The specific error probabilities used in Section VB1.

Input: A group \mathcal{G}_n , a set $\mathcal{A} \subseteq \mathcal{G}_n$, a circuit $\{U_1, \dots, U_\tau\}$, and number of iterations w

Output: A revised circuit $\{G_1^{[w]}, \dots, G_\tau^{[w]}\}$

begin

 Initialize $\{G_1^{[0]}, \dots, G_\tau^{[0]}\} = \{U_1, \dots, U_\tau\}$;

for $\eta \leftarrow 1$ **to** w **do**

 Collect all m neighboring gates $\{(G_1^{[\eta]}, G_2^{[\eta]}), \dots, (G_{m-1}^{[\eta]}, G_m^{[\eta]})\}$ with at least one replaceable candidate neighboring gates $\{G_w^{[\eta+1]} \in \mathcal{A}, G_{w+1}^{[\eta+1]} \in \mathcal{A}\}$;

 Choose $I \in [m-1]$ uniformly at random;

 Replace $(G_I^{[\eta]}, G_{I+1}^{[\eta]})$ by any gate pair in $\{(\tilde{G}_1, \tilde{G}_2) \in \mathcal{A}^2 \mid G_I^{[\eta]} G_{I+1}^{[\eta]} = \tilde{G}_1 \tilde{G}_2\}$ uniformly at random and then obtain the new circuit J ;

 Choose $X \in [0, 1]$ uniformly at random;

if $X \leq \alpha_{G^{[\eta]}, J}(T_\eta)$ **then**

 Set $G^{[\eta+1]} = J$;

else

 Set $G^{[\eta+1]} = G^{[\eta]}$;

end

end

end

Algorithm 2: Pseudo-code for gate-limited simulated annealing.

