Non-Clifford Gates Are Required for Long-Term Memory

Jon Nelson[1]
Joint Center for Quantum Information & Computer Science, University of Maryland, College Park, MD, USA
NIST, Gaithersburg, MD, USA

Joel Rajakumar[1]
Joint Center for Quantum Information & Computer Science, University of Maryland, College Park, MD, USA
NIST, Gaithersburg, MD, USA
IBM T. J. Watson Research Center, Yorktown Heights, NY, USA

Michael J. Gullans
Joint Center for Quantum Information & Computer Science, University of Maryland, College Park, MD, USA
National Institute of Standards and Technology, Gaithersburg, MD, USA

Abstract

We show that all Clifford circuits under interspersed depolarizing noise lose memory of their input exponentially quickly in the depth of the circuit, even when given access to a supply of fresh qubits initialized in arbitrary states. This result applies only in the absence of intermediate measurements, i.e. without intervention by a noiseless external observer such as a classical computer. Nonetheless, this result is surprising given that non-Clifford circuits with access to a supply of fresh qubits can preserve quantum states for arbitrary lengths of time, and even perform fault tolerant quantum computation on them, without requiring any intermediate measurements (Aharonov et al., STOC 1997). Our result shows that such fault tolerance protocols are impossible using only Clifford gates, demonstrating that non-Clifford gates are fundamentally required to store quantum or classical information for long periods of time.

2012 ACM Subject Classification Theory of computation → Quantum information theory; Theory of computation → Quantum complexity theory

Keywords and phrases Fault Tolerance, Quantum Error Correction

Digital Object Identifier 10.4230/LIPIcs.CCC.2026.38

Funding This material is based upon work supported by the U.S. Department of Energy, Office of Science, Accelerated Research in Quantum Computing, Fundamental Algorithmic Research toward Quantum Utility (FAR-Qu). Additional support is acknowledged from IBM, where JR perfomed a portion of this research as an intern. This material is based upon work supported in part by the NSF QLCI award OMA2120757. This work was performed in part at the Kavli Institute for Theoretical Physics (KITP), which is supported by grant NSF PHY-2309135. JN is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2236417.

Acknowledgements We thank Abhinav Deshpande, Kunal sharma, Oles Shtanko, Dominik Hangleiter, Zhi-Yuan Wei, Daniel Malz and Alexey Gorshkov for helpful discussions.

[1] These authors contributed equally to this work.

41st Computational Complexity Conference (CCC 2026).
Editor: Dana Moshkovitz; Article No. 38; pp. 38:1–38:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

It has been known for some time that noisy quantum circuits can fault tolerantly perform arbitrary quantum computations when given access to a constant supply of fresh qubits – or equivalently, the ability to reset a qubit at any time to a fixed state (the *reset gate*) [3]. A weaker task (than computation) that can be made fault tolerant is the implementation of a long-term memory, i.e. storing quantum or classical information in the form of encoded states for long periods of time. Note that the protocol of [3] trivially achieves this task by fault tolerantly implementing the logical identity operation. On the flip side, it has been shown that for noisy quantum circuits without non-unital operations like the reset gate, output states converge towards the maximally mixed state over time [36, 2][2], after which the input state is information-theoretically irrecoverable thus ruling out long-term memory. Here, we show that non-Cliffordness holds a similar power to the reset gate in terms of its necessity for fault-tolerance. In particular, removing the non-Clifford gates from a noisy quantum (or classical) circuit that possesses reset gates also causes it to lose memory of its input exponentially quickly.

We note crucially that we disallow intermediate measurements in our circuit models, which is an identical setup to the threshold theorem proven in [3][3]. This setting is practically relevant as it captures the limitations of near-term quantum devices, where intermediate measurements remain challenging [12]. These restrictions also address the question of whether noisy quantum circuits can protect quantum information *on their own* without the intervention of a noiseless external observer – this has motivated the field of self-correcting quantum memory [18, 5] and is a primary motivation in several fault-tolerance results [2, 3, 7]. Moreover, noiseless classical resources trivially allow the long-term memory of *classical states*, and further, Clifford circuits can store *stabilizer states* when provided with intermediate measurements and noiseless classical processing of measurement outcomes (e.g. syndrome decoding)[23]. In our work, we explicitly disallow noiseless classical resources, which allows us to achieve the strong result that all Clifford protocols lose memory of *any* input state in the presence of noise.

Several recent results have demonstrated a similar loss of memory in noisy *random* quantum circuits that contain resets and other non-unital operations [35, 6, 34, 38]. However, these results differ from ours in a primary aspect, even apart from the assumption of randomness: they show that the input state has negligible effect on the expected value of any *fixed observable* after high enough depth, whereas we show that the input state has negligible effect on all possible observables at once, i.e. that there is no observable that can distinguish two different input states. This is because we prove a convergence in *trace distance*, which, to our knowledge, has never been shown before in a setting that includes non-unital operations and noise of any constant rate. Thus, we anticipate that our analysis techniques may be of broad interest. We state our main theorem below,

▶ **Theorem 1.** *Let Φ denote the channel produced by a circuit of d layers of Clifford gates and reset gates on n qubits, with single-qubit depolarizing noise of strength γ on each qubit after each layer. Let ρ and σ be two arbitrary input states. When $d > d^*$ where $d^* = O(\gamma^{-1} \log(n) \log(n/\epsilon))$,*

$$\|\Phi(\rho) - \Phi(\sigma)\|_1 \leq \epsilon \tag{1}$$

[2] Here, we assume depolarizing noise of any constant rate, on each qubit, after each time step

[3] Note that intermediate measurements can be coherently simulated using the deferred measurement principle, but then any classical processing of the measurement outcomes must be noisy, reversible, and, in our setting, Clifford.

Operationally, our result implies that any two input states become information-theoretically indistinguishable after polylogarithmically many layers of the circuit in the sense that any distinguisher correctly guesses which state it possesses with probability at most $1/2 + o(1/\operatorname{poly}(n))$. This rules out long-term memory in such circuits, since they are incapable of storing quantum or classical information in a manner that can be efficiently extracted (in terms of number of copies) after large time scales. Again, we contrast this with the result of [3], which enables two encoded states to remain efficiently distinguishable for time that scales exponentially with system size[4].

Besides the assumption that the circuit contains only Clifford gates and qubit resets, our result makes no additional assumptions on the circuit. In other words, the statement holds for *any* Clifford circuit using mid-circuit resets in *any* desired locations. In fact, the gates are not required to be geometrically local or even $O(1)$-local. The only requirement is that after each layer of resets and gates a round of constant-strength single-qubit depolarizing channels is applied to each qubit. Finally, the convergence in trace distance occurs for *any* two quantum states. Despite the generality of the result, our proof is relatively short, relying primarily on a counting argument.

1.1 Implications

Our result is somewhat surprising since many error-correction gadgets can be constructed using Clifford gates. For example, one of the most popular types of error correcting codes is a stabilizer code which uses Clifford circuits for both its fault-tolerant encoding and fault-tolerant syndrome extraction protocols [24]. Clifford circuits are also desirable since they propagate Pauli errors nicely making it easier to track errors throughout the circuit, a property which has led to several fault-tolerant implementations of noisy logical Clifford circuits [17, 15, 8, 11]. Although these properties seem useful for correcting errors, our results point out that they are not so powerful on their own, and one must introduce a non-Clifford gate at some point of the computation in order to achieve long-term memory. For example, in order to make intermediate syndrome measurements robust to errors it is often required to perform a majority vote, which cannot be done using Clifford gates. In this way, our result highlights the necessary ingredients for effective fault-tolerant protocols.

Another reason our result is surprising is that magic is often discussed as a computational resource rather than as an error-correction resource. This is because Clifford circuits can be classically simulated efficiently [22] even when augmented with a small number of non-Clifford gates [1, 14]. Thus, a high degree of non-Cliffordness is essential for achieving quantum advantage. In this work, we do not consider the problem of computational hardness but instead focus on the ability to preserve quantum states in memory. This is quite a different task and on the surface does not seem to require departure from Cliffordness as a necessary ingredient. However, our result shows that non-Clifford gates are crucial for error correction in addition to their importance in quantum advantage.

A final feature of our result is that the intermediate reset gates are allowed to produce fresh magic states. One might think that this would allow for the implementation of non-Clifford gates through magic state injection. However, this technique requires a classically controlled Clifford operation as the final step, which is itself a non-Clifford gate. Thus, it is not clear how to use fresh magic states to overcome the limitations of the Clifford circuit. In fact, our no-go result shows that Clifford gates alone cannot exploit the resource

[4] This can be seen by noting that the space and depth overheads of this protocol are polylogarithmic.

of magic states for fault-tolerance, highlighting a distinction between the utility of magic states and non-Clifford gates. This contrasts with the case of computational hardness, where the inclusion of magic states is sufficient to promote Clifford circuits to a model that is hard to sample from classically [13, 45].

1.2 Proof Strategy

Many related noisy circuit problems can be formulated in the Pauli path framework. Broadly speaking, a Pauli path is a trajectory in the Feynman path integral of the circuit with respect to the Pauli basis. In our case, to prove the loss of memory it is sufficient to show that the total contribution of each Pauli path connecting the input to the output state is negligible. To do this, it will be crucial to leverage the property that depolarizing noise causes the contribution of a Pauli path to decay exponentially in the number of non-identities in the path, also called the "weight". This property is the key to many prior works that have also studied noise dynamics in quantum circuits [4, 16, 38, 21, 27].

However, reset gates present a challenge for directly adapting these techniques as they introduce new Pauli paths throughout the circuit. This directly thwarts the Pauli path decay approach, which attempts to suppress the contribution of these paths. To apply Pauli path techniques to circuits with non-unital channels, we use a similar technique to [35, 38, 6, 34], which considers the adjoint map of the noisy circuit. The reason for this is that the adjoint map has the nice property that it is once again unital, and thus the Pauli path decay techniques can be employed on the adjoint map. Unitalness is important for these decay techniques since by definition it means that the map has identity as a fixed point and the presence of depolarizing noise causes these paths to decay until the fixed point is reached.

Even after reformulating our problem as a Pauli path decay problem, the difficulty of the reset gate still presents itself in other subtle ways that require careful analysis. In particular, the adjoint map of the $|0\rangle$-reset gate maps Z to $\mathbb{I}$, which can lower the Pauli weight. This makes the Pauli path less susceptible to noise decay. To address these points, we first treat the depolarization noise as stochastically applying a complete depolarizing error or leaving the qubit unchanged, where a depolarization error has the effect of mapping non-identity Pauli operators to zero. We then reduce our problem to bounding the probability that any non-identity Pauli operator remains at the end of the adjoint circuit. This stochasticization technique was also used for Clifford circuits in [37], but here, we apply it to the adjoint map of the circuit instead. Note that the action of the adjoint reset can now actually help decay the total number of remaining non-identity Pauli operators since it can map two Pauli operators to one (e.g. it maps both $\mathbb{I}$ and Z to $\mathbb{I}$). This property that the number of remaining Pauli operators can only decrease throughout the adjoint circuit is crucial to our analysis.

In our proof, we consider the adjoint circuit in batches of d' layers at a time. Our proof proceeds inductively by alternating two steps. First, we show that all Pauli paths above a given weight w_0 are mapped to zero by a batch of d' layers with high probability. Then we bound the number of remaining Pauli operators by using the property that each input Pauli operator to the batch of layers is mapped to at most one Pauli operator at the output. We then iterate these steps setting $w_1 = w_0/4$ and so on. It turns out that in order to achieve a sufficient decay, it is required that $d' = O(\log n)$ assuming a constant noise strength. Since we must repeat for $O(\log n)$ rounds to decay all possible Pauli weights, we get that no Pauli operators remain at the end of the adjoint circuit after $d = O(\log^2 n)$ total layers.

1.3 Open Questions

Prior work has shown that when circuits have no mechanism for expelling entropy, this entropy accumulates and eventually consumes the circuit [2, 36]. In our work, we analyze the effects of noise in the setting where the circuit is able to pump entropy out of the system by utilizing reset gates. It would be interesting to consider other mechanisms in which the circuit could reduce its entropy. The closest next step would be to generalize our results to show that all Clifford circuits with amplitude damping noise cannot have long-term memory. Similar to the case considered in this paper, there are known fault-tolerant protocols using non-Clifford gates to preserve quantum information under amplitude damping noise [39, 10]. One reason why this could be a tractable problem to solve is that amplitude damping noise acts in a similar manner to a "random reset" error, albeit with some subtle differences. On the other hand, amplitude damping noise also has a unique feature in that it can act as a coherent, but noisy, rotation on stabilizer states [41]; however, it is unclear whether this mechanism can be used as a mid-circuit non-Clifford gate.

Another promising direction is to show that noisy monitored random circuits also have short-term memory. It is known that in the noiseless case, these circuits exhibit a measurement-induced phase transition [29, 40, 28] where above a critical measurement rate, the output of the circuit converges to a single pure state regardless of the input state. On the other hand, below this measurement rate, the circuit preserves an extensive logical codespace for exponentially long times [25]. When noise is introduced, it has been argued using a heuristic mapping of the circuit to a classical Ising model that this phase transition disappears and the circuit loses memory for any constant noise strength and measurement rate [9, 19, 26, 30, 31, 30, 33, 32]. It would be interesting to determine whether our techniques can be used to give a rigorous proof of this statement for the case of random Clifford circuits. Notice that the assumption of randomness is now required since there are known fault-tolerant quantum memory protocols using Clifford gates and intermediate measurements.

Additionally, it would be desirable to lift any of these statements to the case of Haar random circuits. In particular, showing that Haar random circuits under amplitude damping noise have short-term memory would strengthen the result of [35] to the case of trace distance rather than observable estimation which would be a significant improvement. We also point out concurrent work [42], which gives further evidence for this conjecture. Our hope is that the equivalence of second-moment quantities between Clifford circuits and Haar-random circuits may provide one avenue to make progress.

Finally, it is an interesting future direction to consider whether our no-go results on fault-tolerant quantum memory can be lifted to no-go results on computational hardness in certain noise regimes. For example, can we use these techniques to classically simulate Clifford-magic circuits with non-unital noise, similar to [37]? If we can also extend our results to Haar-random circuits as suggested earlier, we may be able to make progress towards classically simulating noisy Haar-random circuits with non-unital noise, which currently presents a significant gap in the literature [20].

2 Background and Notation

2.1 Pauli and Clifford Group

We use $\mathbb{I}$ to denote the identity operator, where the dimension can be inferred by context, and Id to denote the identity channel. The Pauli group is defined as:

$$\mathsf{P}_n := \{1, i, -1, -i\} \times \{\mathbb{I}, X, Y, Z\}^{\otimes n}$$

We define the phaseless Pauli group $\hat{\mathsf{P}}_n \leqslant \mathsf{P}_n$ as:

$$\hat{\mathsf{P}}_n := \{\mathbb{I}, X, Y, Z\}^{\otimes n}$$

We let $|O|$ denote the weight of the Pauli operator $p(O)$, i.e. the number of non-identity qubits. For two sets A and B, we will use $A\backslash B$ to denote $A - A \cap B$.

The Clifford group is defined as:

$$\mathsf{C}_n := \{U \in \mathsf{U}(2^n) : UPU^\dagger \in \mathsf{P}_n \ \forall P \in \mathsf{P}_n\}$$

where $\mathsf{U}(2^n)$ is the group of $2^n \times 2^n$ unitary matrices.

2.2 Stochastic Noise

We define the single-qubit depolarizing noise channel with noise strength γ as:

$$\mathcal{N}_\gamma(\rho) = (1-\gamma)\rho + \gamma\frac{\mathbb{I}}{2}\operatorname{Tr}\rho \tag{2}$$

This channel can be viewed as stochastically applying a complete depolarization error with probability γ, which traces out the given qubit and replaces it with the maximally mixed state. If no error occurs then the channel does nothing. Denoting this depolarization error as $\mathcal{D}$, we have that $\mathcal{N}_\gamma := (1-\gamma)\mathrm{Id} + \gamma\mathcal{D}$. For a given noisy circuit Φ with mid-circuit depolarizing channels, let b be a bitstring that contains a bit for each depolarizing channel representing whether or not $\mathcal{D}$ is applied. We then let Φ_b represent the circuit where each noise channel is replaced with either Id or $\mathcal{D}$ as specified by b. Sampling b according to a binomial distribution parameterized by γ, i.e. $b \sim Bin(nd, \gamma)$, we have that by definition

$$\Phi = \mathbf{E}_b \Phi_b \tag{3}$$

2.3 Reset Channel

We define the qubit reset channel to an arbitrary state ρ as follows:

$$\mathcal{R}_\rho(\sigma) = \rho \operatorname{Tr}\sigma \tag{4}$$

2.4 Adjoint Maps

In our analysis, it will be convenient to use the adjoint of the noisy circuit. For a given channel $\mathcal{N}$, the adjoint map $\mathcal{N}^\dagger$ is defined as follows [43, 44].

$$\langle X, \mathcal{N}(Y)\rangle = \langle \mathcal{N}^\dagger(X), Y\rangle \tag{5}$$

where $\langle X, Y\rangle := \operatorname{Tr} X^\dagger Y$ and X and Y are linear operators on the channel's Hilbert space. From this definition, it can be shown that the adjoint of a given channel $\mathcal{N}(\rho) = \sum_i K_i \rho K_i^\dagger$ is represented by $\mathcal{N}^\dagger(\rho) = \sum_i K_i^\dagger \rho K_i$. Furthermore, given the Pauli transfer matrix of a channel, $T(\mathcal{N})$, which is a matrix with entries $T_{i,j}(\mathcal{N}) = 2^{-n}\langle P_i, \mathcal{N}(P_j)\rangle$ for $P_i, P_j \in \hat{\mathsf{P}}_n$, it is also true that $T(\mathcal{N}^\dagger) = T(\mathcal{N})^\top$.

It is next helpful to explicitly characterize the adjoint map for completely depolarizing errors and qubit resets. First, a depolarizing error $\mathcal{D}$ maps the Pauli matrices as follows

$$\begin{aligned}\mathcal{D}(\mathbb{I}) &= \mathbb{I}\\ \mathcal{D}(X) &= 0\\ \mathcal{D}(Y) &= 0\\ \mathcal{D}(Z) &= 0\end{aligned} \tag{6}$$

Taking the transpose of the associated Pauli transfer matrix we have that $\mathcal{D}^\dagger = \mathcal{D}$.

Next the qubit reset channel maps the Pauli matrices as:

$$\begin{aligned}\mathcal{R}_\rho(\mathbb{I}) &= 2\rho \\ \mathcal{R}_\rho(X) &= 0 \\ \mathcal{R}_\rho(Y) &= 0 \\ \mathcal{R}_\rho(Z) &= 0 \end{aligned} \tag{7}$$

For the adjoint of the reset channel, we can similarly take the transpose of the Pauli matrix associated with Equation (7):

$$\mathcal{R}_\rho^\dagger(\mathbb{I}) = \mathbb{I} \tag{8}$$
$$\mathcal{R}_\rho^\dagger(X) = \mathrm{Tr}(\rho X)\mathbb{I} \tag{9}$$
$$\mathcal{R}_\rho^\dagger(Y) = \mathrm{Tr}(\rho Y)\mathbb{I} \tag{10}$$
$$\mathcal{R}_\rho^\dagger(Z) = \mathrm{Tr}(\rho Z)\mathbb{I} \tag{11}$$

Finally, note that the adjoint of each Clifford gate is also a Clifford gate.

3 Analysis

In general, we will use $A \propto B$ for matrices A, B to refer to the condition that $\exists c \in \mathbb{R} : A = cB$ (note that this condition is satisfied even if $A = 0$). The primary property that we will use in our analysis is that $\Phi_b^\dagger$ is *many-to-one* on the set of Pauli operators since each input Pauli is mapped to at most one output Pauli, as stated below,

▶ **Observation 2.** *For Φ as in Theorem 1 and bitstring b as defined in Section 2.2, for all $s_0 \in \hat{\mathsf{P}}_n$, there exists some $s_1 \in \hat{\mathsf{P}}_n$ such that $\Phi_b^\dagger(s_0) \propto s_1$*

Note that this condition that fails when considering non-Clifford gates, whose adjoint maps can take one Pauli operator to many. Our techniques apply generally to all noisy circuits that can be represented as a mixture over channels whose adjoint maps have this many-to-one property.

3.1 Reduction to Pauli Path Survival Probability

We begin by analyzing our desired quantity in the Pauli basis, which allows us to reduce our problem to bounding the probability that a non-identity Pauli operator is output by the adjoint of the circuit. This is captured in the following lemma,

▶ **Lemma 3.** *For $\Phi, \rho, \sigma, \gamma$ as in Theorem 1,*

$$\|\Phi(\rho) - \Phi(\sigma)\|_1 \leq 2\Pr(\exists s \in \hat{\mathsf{P}}_n : \Phi_b^\dagger(s) \not\propto \mathbb{I}) \tag{12}$$

where the probability $\Pr(\cdot)$ is taken over bitstrings $b \sim Bin(nd, \gamma)$, which specify the locations of depolarization errors in Φ_b.

Proof. For any error configuration b, note the following fact.

$$\Phi_b(\rho) = \Phi_b(\sigma) \tag{13}$$
$$\iff \forall s \in \hat{\mathsf{P}}_n : \mathrm{Tr}(\Phi_b(\rho)s) = \mathrm{Tr}(\Phi_b(\sigma)s) \tag{14}$$
$$\iff \forall s \in \hat{\mathsf{P}}_n : \mathrm{Tr}(\rho\Phi_b^\dagger(s)) = \mathrm{Tr}(\sigma\Phi_b^\dagger(s)) \tag{15}$$

The above property follows from the fact that all density matrices have a unique decomposition in the Pauli basis. Now,

$$\begin{aligned}
\|\Phi(\rho) - \Phi(\sigma)\|_1 &= \|\mathbf{E}_b \Phi_b(\rho) - \Phi_b(\sigma)\|_1 && (16)\\
&\leq \mathbf{E}_b \|\Phi_b(\rho) - \Phi_b(\sigma)\|_1 && (17)\\
&\leq \Pr(\Phi_b(\rho) = \Phi_b(\sigma)) \cdot 0 + \Pr(\Phi_b(\rho) \neq \Phi_b(\sigma)) \cdot 2 && (18)\\
&= 2 \Pr(\Phi_b(\rho) \neq \Phi_b(\sigma)) && (19)\\
&= 2 \Pr(\exists s \in \hat{\mathsf{P}}_n : \mathrm{Tr}(\rho \Phi_b^\dagger(s)) \neq \mathrm{Tr}(\sigma \Phi_b^\dagger(s))) && (20)\\
&\leq 2 \Pr(\exists s \in \hat{\mathsf{P}}_n : \Phi_b^\dagger(s) \propto \mathbb{I}) && (21)
\end{aligned}$$

In the last step, notice that if $\Phi_b^\dagger(s) = c\mathbb{I}$ for some $c \in \mathbb{R}$ then $\mathrm{Tr}(\rho c\mathbb{I}) = \mathrm{Tr}(\sigma c\mathbb{I}) = c$. Therefore, the only way for $\mathrm{Tr}(\rho\Phi_b^\dagger(s)) \neq \mathrm{Tr}(\sigma\Phi_b^\dagger(s))$ is if $\Phi_b^\dagger(s)$ is not proportional to the identity. A crucial step of this proof is Equation (18), where we have used the fact that for any choice of b, the trace distance $\|\Phi_b(\rho) - \Phi_b(\sigma)\|_1$ is always bounded by two. This is true because Φ_b is a quantum channel and the trace distance between two density matrices is at most two. ◀

It now remains to upper bound $\Pr(\exists s \in \hat{\mathsf{P}}_n : \Phi_b^\dagger(s) \propto \mathbb{I})$.

3.2 Grouping Pauli Operators Based on Weight

Our argument will rely on a careful accounting of the non-identity Pauli operators that remain at various points in the adjoint of the circuit. The probability that a Pauli operator is mapped to zero is closely related to its weight since the depolarizing channel maps non-identity Pauli operators to zero with constant probability. For a circuit Φ, it is useful to divide a given set of Pauli operators G into subsets based on their minimum weight throughout the circuit. These subsets are formally defined as follows

▶ **Definition 4.** *For Φ as in Theorem 1, let $S_w(\Phi, G) \subseteq \hat{\mathsf{P}}_n$ be the set:*

$$S_w(\Phi, G) := \{s \in G : \min_i |\Phi_{i\leftarrow}(s)| = w\}. \tag{22}$$

where $\Phi_{i\leftarrow}$ denotes the first i layers of the circuit. Here, each layer constitutes a round of gates, a round of depolarizing channels (or sampled depolarizing errors) and a round of arbitrarily placed qubit resets. When G is not defined we take it to be $\hat{\mathsf{P}}_n$. We define $S_{\geq w}(\Phi, G)$ and $S_{<w}(\Phi, G)$ in a similar manner.

A key fact we will use is the following,

▶ **Observation 5.** *For Φ, d, γ as in Theorem 1, for all $s \in S_w(\Phi, G)$, $\Pr(\Phi_b(s) \neq 0) \leq (1-\gamma)^{wd}$*

This is simply due to the fact that each layer of depolarizing channels will encounter at least w non-identity qubits since this is the minimum weight throughout each layer of the circuit. The Pauli operator survives each depolarizing channel acting on a non-identity with probability $1 - \gamma$. Because there are d layers, this must occur at least wd times, and so the probability of surviving all depolarizing errors is $(1-\gamma)^{wd}$.

3.3 Batched Counting Argument

Our strategy will be to consider $d' = O(\gamma^{-1} \log n)$ layers at a time and show that with high probability the number of remaining Pauli operators decreases substantially after each batch of layers.

▶ **Lemma 6.** *For Φ, d, γ as in Theorem 1, if $d > d^*$ where $d^* = O(\gamma^{-1} \log(n) \log(n/\epsilon))$, then*

$$\Pr(\exists s \in \hat{\mathsf{P}}_n : \Phi_b^\dagger(s) \not\propto \mathbb{I}) \leq \epsilon \tag{23}$$

where the probability $\Pr(\cdot)$ is taken over bitstrings $b \sim Bin(nd, \gamma)$, which specify the locations of depolarization errors in Φ_b.

Note that Theorem 1 is implied by the combination of Lemma 3 and Lemma 6.

Proof. First, notice that we do not need to differentiate between two Pauli operators with different real-valued coefficients at any point in the adjoint map of the circuit. This is because, the coefficient will not affect whether or not this operator gets mapped to zero or identity by the adjoint map of the circuit. To handle this, we introduce the following notation: for any operator $O \propto P$ where $P \in \hat{\mathsf{P}}_n$, we let $p(O) := P$.

Our proof proceeds by first breaking up the adjoint of the circuit into batches of d' layers. Let $\Phi_{b,j}^\dagger$ be the jth batch of d' layers of the adjoint of the circuit. For each j, we will associate a weight limit w_j, where $w_j = \lceil w_{j-1}/4 \rceil$ and $w_0 = n$. Define the sequences of sets $F_0, \dots F_{j-1}, F_j$ and $G_0, \dots G_{j-1}, G_j$ inductively as follows,

$$F_0 = G_0 = \{\hat{\mathsf{P}}_n\} \tag{24}$$

$$F_j = \{p(\Phi_{b,j}^\dagger(s)) \ \forall s \in F_{j-1} \text{ s.t. } \Phi_{b,j}^\dagger(s) \neq 0\} \tag{25}$$

$$G_j = \{p(\Phi_{b,j}^\dagger(s)) \ \forall s \in S_{<w_j}(\Phi_{b,j}^\dagger, G_{j-1}) \text{ s.t. } \Phi_{b,j}^\dagger(s) \neq 0\} \tag{26}$$

In other words, F_j and G_j are defined as the sets of all unscaled output Pauli operators of $\Phi_{b,j}^\dagger$ starting from the input Pauli operators in F_{j-1} and G_{j-1} respectively, where G_j has the added restriction that the minimum weight of the operator as it passes through $\Phi_{b,j}^\dagger$ is less than w_j. With this notation, our new goal is to prove that $\Pr(F_j \neq \{\mathbb{I}\}) \leq \epsilon$ for some large enough j.

We will iterate through the batches while computing probabilities over $b \sim Bin(nd, \gamma)$ as follows: when we are considering the jth batch of the circuit, we will fix all locations of depolarizing errors in batches $1 \dots j-1$, and use $\Pr_j(\cdot)$ to to denote the probability over locations of depolarizing errors within $\Phi_{b,j}^\dagger$. Our inductive hypothesis is that $\Pr_{j-1}(F_{j-1} \neq G_{j-1}) \leq (j-1)\epsilon'$ and that $|G_{j-1}| \leq d'(9n)^{w_{j-1}}$. This is clearly satisfied in the base case of $j - 1 = 0$. Our inductive step will be to show that $\Pr_j(F_j \neq G_j) \leq j'\epsilon$, and that $|G_j| \leq d'(9n)^{w_j}$. We will iterate this argument until $w_j = 1$, which occurs at $j = O(\log n)$, at which point $G_j = \{\mathbb{I}\}$ since all operators in G_j must have weight less than w_j[5]. At this point, by induction, $\Pr(F_j \neq \{\mathbb{I}\}) \leq O(\epsilon' \log n)$, and so we choose $\epsilon' = O(\epsilon/\log n)$ to complete the proof.

[5] Note that the bounds $|G_j| \leq d'(9n)^{w_j}$ never reaches 1, and we have established $|G_j| = 1$ via a different argument

First, we upper bound $\Pr(F_j \neq G_j)$,

$$\begin{aligned}
\Pr_j(F_j \neq G_j) &= \Pr_j(F_j \neq G_j | F_{j-1} = G_{j-1}) \Pr_{j-1}(F_{j-1} = G_{j-1}) && (27)\\
&+ \Pr_j(F_j \neq G_j | F_{j-1} \neq G_{j-1}) \Pr_{j-1}(F_{j-1} \neq G_{j-1}) && (28)\\
&\leq \Pr_j(F_j \neq G_j | F_{j-1} = G_{j-1}) + (j-1)\epsilon' && \text{(inductive hypothesis)}\\
&\leq \Pr_j(\exists s \in S_{\geq w_j}(\Phi^\dagger_{b,j}, G_{j-1}) : \Phi^\dagger_{b,j}(s) \neq 0) + (j-1)\epsilon' && (29)\\
&\leq \sum_{s \in S_{\geq w_j}(\Phi^\dagger_{b,j}, G_{j-1})} \Pr_j(\Phi^\dagger_{b,j}(s) \neq 0) + (j-1)\epsilon' && \text{(union bound)}\\
&\leq |G_{j-1}|(1-\gamma)^{w_j d'} + (j-1)\epsilon' && (30)\\
&\leq d'(9n)^{w_{j-1}}(1-\gamma)^{w_{j-1}d'/4} + (j-1)\epsilon' && \text{(inductive hypothesis)}\\
&\leq \epsilon' + (j-1)\epsilon' && (\text{setting } d' = O(\gamma^{-1}\log(n/\epsilon')))\\
&= j\epsilon' && (31)
\end{aligned}$$

Next, we upper bound $|G_j|$. Consider every possible *intermediate* Pauli operator at a given layer of $\Phi^\dagger_{b,j}$, whose weight is less than w_j. We can exploit the property that each such Pauli operator, whichever layer it is in, gets mapped by $\Phi^\dagger_{b,j}$ to at most one output Pauli operator, which may or may not be in G_j. Therefore, we can upper bound $|G_j|$ using the size of this set. There are d' layers, w_j possible values of the minimum weight, at most $\binom{n}{w_j}$ ways to choose the locations of its non-identities [6], and at most 3^{w_j} ways to assign non-identities to X, Y, and Z. So we have,

$$\begin{aligned}
|G_j| &\leq w_j d' \binom{n}{w_j} 3^{w_j} && (32)\\
&\leq w_j d' (\frac{en}{w_j})^{w_j} 3^{w_j} && (33)\\
&\leq d'(9n)^{w_j} && (34)
\end{aligned}$$

◀

References

1 Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. A*, 70:052328, November 2004. doi:10.1103/PhysRevA.70.052328.

2 D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan. Limitations of noisy reversible computation, 1996. arXiv:quant-ph/9611028.

3 Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error rate. *SIAM Journal on Computing*, 38(4):1207–1282, 2008. doi:10.1137/S0097539799359385.

4 Dorit Aharonov, Xun Gao, Zeph Landau, Yunchao Liu, and Umesh Vazirani. A polynomial-time classical algorithm for noisy random circuit sampling. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 945–957, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3564246.3585234.

5 R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki. On thermal stability of topological qubit in kitaev's 4d model. *Open Systems & Information Dynamics*, 17(01):1–20, 2010. doi:10.1142/S1230161210000023.

[6] assuming $w_j \leq n/2$, which is valid since only $w_0 > n/2$ and this is the base case

6 Armando Angrisani, Antonio A. Mele, Manuel S. Rudolph, M. Cerezo, and Zoe Holmes. Simulating quantum circuits with arbitrary local noise using pauli propagation, 2025. arXiv:2501.13101.

7 A. Anshu, Y. Liu, Q. Nguyen, and C. Pattison, 2025. Simons talk available at: https://simons.berkeley.edu/talks/quynh-nguyen-harvard-university-2025-05-28.

8 Dave Bacon, Steven T. Flammia, Aram W. Harrow, and Jonathan Shi. Sparse quantum codes from quantum circuits. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 327–334, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2746539.2746608.

9 Yimu Bao, Soonwon Choi, and Ehud Altman. Symmetry enriched phases of quantum circuits. *Annals of Physics*, 435:168618, 2021. doi:10.1016/j.aop.2021.168618.

10 Michael Ben-Or, Daniel Gottesman, and Avinatan Hassidim. Quantum refrigerator, 2013. arXiv:1301.1995.

11 Thiago Bergamaschi and Yunchao Liu. On Fault Tolerant Single-Shot Logical State Preparation and Robust Long-Range Entanglement. In Raghu Meka, editor, *16th Innovations in Theoretical Computer Science Conference (ITCS 2025)*, volume 325 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 16:1–16:9, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2025.16.

12 Dolev Bluvstein, Simon J. Evered, Alexandra A. Geim, Sophie H. Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, J. Pablo Bonilla Ataides, Nishad Maskara, Iris Cong, Xun Gao, Pedro Sales Rodriguez, Thomas Karolyshyn, Giulia Semeghini, Michael J. Gullans, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, 2024.

13 Adam Bouland, Joseph F. Fitzsimons, and Dax Enshan Koh. Complexity Classification of Conjugated Clifford Circuits. In Rocco A. Servedio, editor, *33rd Computational Complexity Conference (CCC 2018)*, volume 102 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:25, Dagstuhl, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2018.21.

14 Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Phys. Rev. Lett.*, 116:250501, June 2016. doi:10.1103/PhysRevLett.116.250501.

15 Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. Quantum advantage with noisy shallow circuits. *Nature Physics*, 16(10):1040–1045, July 2020. doi:10.1038/s41567-020-0948-z.

16 Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum*, 1:8, April 2017. doi:10.22331/q-2017-04-25-8.

17 Nicolas Delfosse and Adam Paetznick. Spacetime codes of clifford circuits, 2023. doi:10.48550/arXiv.2304.05943.

18 Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill. Topological quantum memory. *Journal of Mathematical Physics*, 43(9):4452–4505, September 2002. doi:10.1063/1.1499754.

19 Beatriz C. Dias, Domagoj Perkovic, Masudul Haque, Pedro Ribeiro, and Paul A. McClarty. Quantum Noise as a Symmetry-Breaking Field. *arXiv preprint*, 2022. arXiv:2208.13861.

20 Bill Fefferman, Soumik Ghosh, Michael Gullans, Kohdai Kuroiwa, and Kunal Sharma. Effect of nonunital noise on random-circuit sampling. *PRX Quantum*, 5:030317, July 2024. doi:10.1103/PRXQuantum.5.030317.

21 Xun Gao and Luming Duan. Efficient classical simulation of noisy quantum computation, 2018. arXiv:1810.03176.

22 Daniel Gottesman. The heisenberg representation of quantum computers, 1998. arXiv:quant-ph/9807006.

23 Daniel Gottesman. An introduction to quantum error correction and fault-tolerant quantum computation, 2009. arXiv:0904.2557.

24 Daniel Gottesman. Fault-tolerant quantum computation with constant overhead. *Quantum Info. Comput.*, 14(15–16):1338–1372, November 2014. doi:10.26421/QIC14.15-16-5.

25 Michael J. Gullans and David A. Huse. Dynamical purification phase transition induced by quantum measurements. *Physical Review X*, 10(4), October 2020. doi:10.1103/physrevx.10.041020.

26 Shao-Kai Jian, Chunxiao Liu, Xiao Chen, Brian Swingle, and Pengfei Zhang. Quantum error as an emergent magnetic field. *arXiv preprint*, 2021. arXiv:2106.09635.

27 Julia Kempe, Oded Regev, Falk Unger, and Ronald de Wolf. Upper bounds on the noise threshold for fault-tolerant quantum computing. In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *Automata, Languages and Programming*, pages 845–856, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. doi:10.1007/978-3-540-70575-8_69.

28 Yaodong Li, Xiao Chen, and Matthew P. A. Fisher. Quantum zeno effect and the many-body entanglement transition. *Phys. Rev. B*, 98:205136, November 2018. doi:10.1103/PhysRevB.98.205136.

29 Yaodong Li, Xiao Chen, and Matthew P. A. Fisher. Measurement-driven entanglement transition in hybrid quantum circuits. *Phys. Rev. B*, 100:134306, October 2019. doi:10.1103/PhysRevB.100.134306.

30 Yue Li and Martin Claassen. Statistical mechanics of monitored dissipative random circuits. *Phys. Rev. B*, 108:104310, September 2023. doi:10.1103/PhysRevB.108.104310.

31 Zhi Li, Shengqi Sang, and Timothy H. Hsieh. Entanglement dynamics of noisy random circuits. *Phys. Rev. B*, 107:014307, January 2023. doi:10.1103/PhysRevB.107.014307.

32 Shuo Liu, Ming-Rui Li, Shi-Xin Zhang, and Shao-Kai Jian. Entanglement structure and information protection in noisy hybrid quantum circuits. *Phys. Rev. Lett.*, 132:240402, June 2024. doi:10.1103/PhysRevLett.132.240402.

33 Shuo Liu, Ming-Rui Li, Shi-Xin Zhang, Shao-Kai Jian, and Hong Yao. Universal kardar-parisi-zhang scaling in noisy hybrid quantum circuits. *Phys. Rev. B*, 107:L201113, May 2023. doi:10.1103/PhysRevB.107.L201113.

34 Victor Martinez, Armando Angrisani, Ekaterina Pankovets, Omar Fawzi, and Daniel Stilck França. Efficient simulation of parametrized quantum circuits under nonunital noise through pauli backpropagation. *Phys. Rev. Lett.*, 134:250602, June 2025. doi:10.1103/j1gg-s6zb.

35 Antonio Anna Mele, Armando Angrisani, Soumik Ghosh, Sumeet Khatri, Jens Eisert, Daniel Stilck França, and Yihui Quek. Noise-induced shallow circuits and absence of barren plateaus, 2024. arXiv:2403.13927.

36 Alexander Müller-Hermes, Daniel Stilck França, and Michael M. Wolf. Relative entropy convergence for depolarizing channels. *Journal of Mathematical Physics*, 57(2), January 2016. doi:10.1063/1.4939560.

37 Jon Nelson, Joel Rajakumar, Dominik Hangleiter, and Michael J. Gullans. *Polynomial-Time Classical Simulation of Noisy Quantum Circuits with Naturally Fault-Tolerant Gates*, pages 1309–1331. SIAM, 2026. doi:10.1137/1.9781611978971.50.

38 Thomas Schuster, Chao Yin, Xun Gao, and Norman Y. Yao. A polynomial-time classical algorithm for noisy quantum circuits, 2024. doi:10.48550/arXiv.2407.12768.

39 Oles Shtanko and Kunal Sharma. Complexity of local quantum circuits under nonunital noise, 2024. arXiv:2411.04819.

40 Brian Skinner, Jonathan Ruhman, and Adam Nahum. Measurement-induced phase transitions in the dynamics of entanglement. *Phys. Rev. X*, 9:031009, July 2019. doi:10.1103/PhysRevX.9.031009.

41 Fabian Ballar Trigueros and José Antonio Marín Guzmán. Nonstabilizerness and error resilience in noisy quantum circuits, 2025. arXiv:2506.18976.

42 Su un Lee, Soumik Ghosh, Changhun Oh, Kyungjoo Noh, Bill Fefferman, and Liang Jiang. Classical simulation of noisy random circuits from exponential decay of correlation, 2025. arXiv:2510.06328.

43 John Watrous. cs.uwaterloo.ca. https://cs.uwaterloo.ca/~watrous/QIT-notes/QIT-notes.pdf. [Accessed 11-09-2025].

44 Mark Wilde. markwilde.com. https://markwilde.com/teaching/2015-fall-qit/lectures/lecture-10.pdf. [Accessed 11-09-2025].

45 Mithuna Yoganathan, Richard Jozsa, and Sergii Strelchuk. Quantum advantage of unitary clifford circuits with magic state inputs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2225):20180427, May 2019. doi:10.1098/rspa.2018.0427.