SAND2021-13613C

Extended Abstract: Asymptotically Tighter Simulation of Universal Quantum Outcomes

Lucas Kocia
Sandia National Laboratories, Livermore, California 94550, U.S.A.

Genele Tulloch
North Carolina Central University

We are rapidly reaching the point with today's quantum computers where we will be unable to verify their outcomes with our current classical computers. Arguably, in some applications, this threshold has already been reached [1]. Soon this may also be the case with fault-tolerant or near-term intermediate-scale quantum (NISQ) computers [2] that implement useful quantum algorithms.

Validating the performance of NISQ devices and establishing a clearly-defined threshold for when they can outperform current computers is accomplished by directly simulating their outcomes on classical computers. The community needs more efficient classical algorithms and they need to know how much more efficient they can get. This last desired quality is often missing in currently available methods. Here, we accomplish both of these goals.

Universal quantum computation requires a universal gateset, appropriate initial states and set of measurements. A common universal set that is widely used consists of stabilizer states, the Clifford+T gateset, and Pauli measurements. For classical computers, simulating everything other than the T gates can be done very efficiently in polynomial time [3]. On the other hand, T gates have been found to be exponentially difficult to simulate.

Fortunately, classical simulation of universal quantum outcomes is made easier somewhat by the fact that quantum computers only produce output up to additive error [4], and so there is *a priori* no reason to simulate their outcomes beyond this. Simulating a given state as it undergoes Clifford and T gate unitary evolution up to some additive error is an L_1 minimization problem, which are known to scale more favorably compared to L_0 or L_2 minimizations appropriate when there are more stringent error constraints. The most natural state decomposition that is used for such classical simulation is the stabilizer state decomposition, because classical computers can calculate their inner products especially efficiently. The corresponding number of these stabilizer states sufficient for additive error decompositions is called the *stabilizer extent*.

Current state-of-the-art classical simulators of quantum outcomes from this gateset scale as $\mathcal{O}(\xi^t\delta^{-2})$ [5–7] and $\mathcal{O}(\xi^t\delta^{-1})$ [8], where $\xi^t \equiv 2^{\sim 0.228t}$ is the stabilizer extent of the T gate magic state and δ is the additive error. Furthermore, it is conjectured [6] that this asymptotic scaling is tight w.r.t. t. The most promising route to find reductions in this algorithm is to look for at the next-order terms.

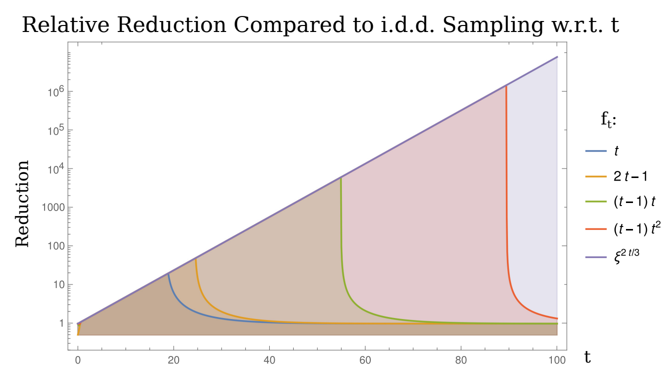

FIG. 1. [From Technical Paper #1] The relative improvement in scaling w.r.t. t produced by correlated sampling with different $0 \le f_t \le \mathcal{O}(\xi^{2t/3})$ compared to the state-of-the-art. In practice, the desired δ determines which f_t you can attain: $\delta^2 \gg (\xi^t - f_t)^{-1} \implies f_t \ll \delta^{-2} - \xi^t$.

Despite their dimunitive name, next-order effects can offer significant improvements for the intermediate (non-asymptotic) t regime, which is precisely the regime that NISQ devices are in. For a sense of this, see Figure 1 for a preview of the relative improvement in scaling that can be attained from including such next-order terms. Unfortunately, existing methods all saturate the asymptotic conjectured lower bound w.r.t. t even when they are not in the asymptotic limit of $t \to \infty$ and it is not immediately clear how to extend them to next order in t.

We take a first crack at this problem in [9] (attached Technical Paper #2) and introduce the tool of correlated L_1 sampling. The main technique that we use is supplementing independent L_1 sampling with correlated samples of "dissimilar" states that lead to a better average. Given an ensemble of M independently sampled states, which have an

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

expectation value close to the desired state we are simulating to d-accuracy, we transform the ensemble to a correlated one by supplementing each independently sampled state in a M/f_t subset of the ensemble that are "dissimilar" to each other.

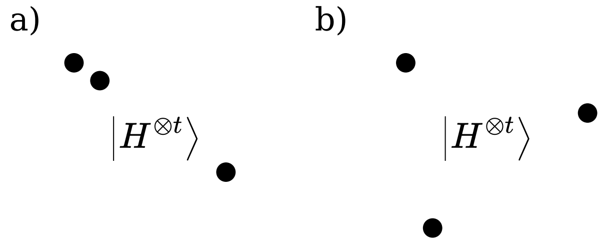

FIG. 2. [From Technical Paper #1] While the expectation value of independent sampling will approach the desired state, any finite set will on average deviate from the desired state. By replacing some independent samples with correlated samples, "spurious" sample deviation from the desired state can be minimized, at the expense of the variance of the distribution. This latter effect is negligible if the distribution is sufficiently peaked already.

The expectation value of this ensemble is closer to the desired state but converges more slowly for the same δ-error. If $\delta \gg (\xi^t - f_t)$, this slower convergence is negligible (i.e. your probability distribution is already sufficiently peaked). A sketch of this phenomenon can be seen in Fig. 2.

Using this technique of correlated sampling produces a linear reduction: $\mathcal{O}((\xi^t - f_t)\delta^2)$ where $f_t = t$. We subsequently extend these results (attached Technical Paper #1) to $f_t = 2t - 1$ states. We then prove that this set of $2t - 1$ states can be used as a "seed" set to generate any higher number of appropriately correlated states up to $f_t = \mathcal{O}(\xi^{2t/3})$. This is an additive exponential improvement over the state-of-the-art!

How much further can such an L_1 technique be pushed? Given widely accepted complexity conjectures, we answer this question by proving the upper bound $f_t \in o(\xi^t)$, thereby proving that our algorithm is tight up to $o(\xi^{t/3})$. Again, see Figure 1 for a plot of the relative improvement in scaling of our new algorithm compared to the state-of-the-art.

This means that weak simulation based on the minimal L_1 stabilizer decomposition of the T gate cannot be further improved beyond an additive factor of $o(\xi^{t/3}$ with respect to t scaling. This settles the question of how much it is possible to push the next-order term in weak simulation to speed up intermediate-size problems. To our knowledge, this work constitutes the first weak simulation algorithm that has lowered this bound's dependence on finite t in the worst-case. It holds great promise in rendering classical simulation of NISQ outcomes more tractable on today's classical computers.

---

[1] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

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

[3] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004.

[4] Hakop Pashayan, Stephen D Bartlett, and David Gross. From estimation of quantum probabilities to simulation of quantum circuits. *Quantum*, 4:223, 2020.

[5] Sergey Bravyi, Graeme Smith, and John A Smolin. Trading classical and quantum computational resources. *Physical Review X*, 6(2):021043, 2016.

[6] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Physical review letters*, 116(25):250501, 2016.

[7] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019.

[8] James R Seddon, Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl T Campbell. Quantifying quantum speedups: improved classical simulation from tighter magic monotones. *arXiv preprint arXiv:2002.06181*, 2020.

[9] Lucas Kocia. Improved weak simulation of universal quantum circuits by correlated l_1 sampling. *arXiv preprint arXiv:2104.07250*, 2021.

—

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under the Accelerated Research in Quantum Computing program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National

Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

This research was supported in part by an appointment to the National Nuclear Security Administration (NNSA) Minority Serving Institutions Internship Program (MSIIP) sponsored by the NNSA and administered by the Oak Ridge Institute for Science and Education (ORISE).

Technical Paper #1 of 2

Asymptotically Tighter Simulation of Universal Quantum Outcomes

Lucas Kocia
Sandia National Laboratories, Livermore, California 94550, U.S.A.

Genele Tulloch
North Carolina Central University

We present a classical algorithm for simulating quantum circuit outcomes consisting of the universal Clifford+T gateset that scales as $\mathcal{O}((\xi^t - f_t)\delta^{-2})$ for $\delta^2 \gg (\xi^t - f_t)^{-1}$, where δ is the additive error, $\xi^t = 2^{\sim 0.228t}$ is the stabilizer extent, and t is the number of T gates. In our algorithm, we demonstrate an exponential increase of $f_t \in \mathcal{O}(\xi^{2t/3})$, compared to the state-of-the-art $f_t = t$. Numerical demonstrations support our claims. Given widely accepted complexity conjectures, we prove the upper bound $f_t \in o(\xi^t)$. As a consequence, classical simulation of quantum outcomes based on minimal L_1 stabilizer decompositions of the T gate magic state cannot be further decreased compared to our algorithm beyond an additive factor of $o(\xi^{t/3})$ with respect to t.

The rapid progress in the practical development of quantum computers holds great promise for producing a quantum device that can solve many key problems of interest that do not scale favorably on classical computers. Small fault-tolerant quantum computers and larger near-term intermediate-scale quantum (NISQ) computers often target the implementation of the popular Clifford+T gateset, with stabilizer state initial states and final Pauli measurements. Validating their performance, as well as establishing a clearly defined threshold for when they are outperforming current computers, requires directly simulating their outcomes on classical computers. Such a simulation scales exponentially with the number of T gates.

Quantum computers only output outcomes that are correct up to an additive error, which greatly simplifies their classical simulation. This is because an additive error constraint reduces the problem to an L_1 minimization, which are known to scale more favorably compared to L_0 or L_2 minimizations that are more relevant to multiplicative error constraints. Given any state decomposed into an orthonormal basis, $|\psi\rangle = \sum_i c_i|\phi_i\rangle$, the L_1 norm of the state is the sum of the absolute values of its coefficients: $\sum_i |c_i|$. The L_1 norm is a multiplicative quantity: the L_1 norm of $|\psi\rangle^{\otimes t}$ is the L_1 norm of $|\psi\rangle$ taken to the power of t.

Here we will be interested in stabilizer decompositions of given states $|\psi\rangle$. Calling the sum of the absolute values of the coefficients of non-orthogonal stabilizer states an L_1 norm is technically an abuse of terminology, since the stabilizer states are overcomplete. Therefore, this quantity is often called the *stabilizer extent*, ξ, in this context.

Nevertheless, the stabilizer extent shares an important property with minimal L_1 norms: it is multiplicative, though for only one-, two-, and three-qubit tensored states. Luckily, this is sufficient for our uses since we will be primarily concerned with one-qubit tensored T gate magic states.

The T gate magic state

$$|T\rangle = \frac{1}{\sqrt{2}}(|0\rangle + \sqrt{i}|1\rangle), \tag{1}$$

is Clifford-isomorphic to the H magic state

$$|H\rangle = e^{-i\pi/8} SH|T\rangle, \tag{2}$$

which can be decomposed into stabilizer states:

$$|H\rangle = \left(2\cos\frac{\pi}{8}\right)^{-1}\left(|\tilde{0}\rangle + |\tilde{1}\rangle\right), \tag{3}$$

where

$$|\tilde{0}\rangle = |0\rangle, \quad \text{and} \quad |\tilde{1}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \tag{4}$$

It is often useful to write $|H\rangle^{\otimes t}$ in terms of its single-qubit basis:

$$|H\rangle^{\otimes t} = \sum_{x \in \mathbb{F}_2^t} |\tilde{x}\rangle. \tag{5}$$

The stabilizer state decomposition given by Eq. 3 is the natural decomposition to use when interested in representing $|H\rangle^{\otimes t}$ only up to some additive error δ, since it has minimal L_1 norm with respect to the stabilizer state basis, i.e. this decomposition saturates the H gate magic state stabilizer extent: $\xi^t = \left(\cos\frac{\pi}{8}\right)^{-t/2} = 2^{\sim 0.228t}$.

Moreover, since we only want to capture $|H\rangle^{\otimes t}$ up to additive error δ, we can *sparsify* our stabilizer state decomposition of $|H\rangle^{\otimes t}$ in the sum given by Eq. 5 to produce an approximation $|\psi\rangle$. We only need to keep enough $|\tilde{i}\rangle$ states such that $||\psi\rangle - |H\rangle^{\otimes t}|_1 \leq \delta$.

Selecting which $|\tilde{x}\rangle$ to keep can be done by viewing the set $\{|\tilde{x}\rangle\}$ as an ensemble defined by an independent and identically distributed (i.i.d.) random variable. Sampling $|T\rangle^{\otimes t}$ (or equivalently $|H\rangle^{\otimes t}$) this way requires $\mathcal{O}(\xi^t\delta^{-2})$ states [1]. Using a slightly different normalization method, sampling can be improved in scaling with δ so that it requires $\mathcal{O}(\xi^t\delta^{-1})$ states [2]. Some improvement has also been made in the next step of estimating outcome probabilities for certain values of these probabilities [3] but still require $\mathcal{O}(\xi^t\delta^{-1})$ states.

Here we will be concerned with reducing the scaling with respect to t. It turns out that, to leading order, the scaling of sampling $|H\rangle^{\otimes t}$ is likely tight and has been conjectured [1] to be lower bounded by the same factor: $\Omega(\xi^t)$. However, the optimal next-order leading term is unknown and can still lead to a substantial reduction in the scaling for intermediate t, of particular interest in simulating NISQ devices. Unfortunately, existing methods all saturate the asymptotic conjectured lower bound w.r.t. t even when they are not in the asymptotic limit of $t \to \infty$ and it is not clear from their derivation how to improve them.

We will show how such next-order reductions in t can be derived and how they greatly increase the size of universal quantum circuits that are simulatable by today's classical computers.

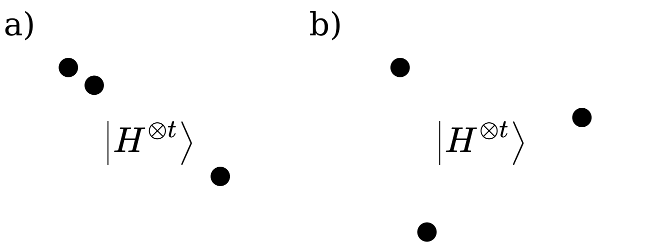

FIG. 1. While the expectation value of independent sampling will approach the desired state, any finite set will on average deviate from the desired state. By replacing some independent samples with correlated samples, "spurious" sample deviation from the desired state can be minimized, at the expense of the variance of the distribution. This latter effect is negligible if the distribution is sufficiently peaked already.

The main technique that we will use is supplementing independent sampling with correlated samples of "dissimilar" states, that thereby lead to a better average. Given an ensemble of M independently sampled states, $\{|\tilde{x}_i\rangle\}_i$, from the stabilizer state space, which have an expectation value close to $|H^{\otimes t}\rangle$, we transform the ensemble to a correlated one by supplementing each independently sampled state in a M/f_t subset of the ensemble, with $f_t = t/2$ bit-flipped stabilizer states—which are therefore "dissimilar" to each other by $t/2$ bitflips—and discard the rest. These supplemental states are maximally dissimilar from each other in that they constitute the optimal largest number of supplemental states with the largest mutual dissimilarity.

The expectation value of this ensemble is closer to $|H^{\otimes t}\rangle$ but converges more slowly for the same δ-error. If $\delta \gg (\xi^t - f_t)$, this slower convergence is negligible (i.e. your probability distribution is already sufficiently peaked; the variance is sufficiently small). A sketch of the phenomenon can be seen in Fig. 1.

This procedure is a modification of the popular SPARSIFY algorithm [4] and is formally proven in [5] (see Theorem 1 therein), where it was shown that

$$\mathbb{E}(\||H^{\otimes t}\rangle - |\psi\rangle\|^2) \leq \frac{\xi_\phi^t}{k} - \frac{\gamma}{k} \leq \delta^2, \tag{6}$$

for

$$\gamma = 1 + f_t - \sum_i^k \sum_j^{f_t} \xi^t \mathbb{E}(\langle \tilde{x}_i | \tilde{y}_{\sigma_i(j)} \rangle), \tag{7}$$

where $\sigma_i(j)$ maps the $\tilde{x}_i$ to the f_t states $\{\tilde{y}_{\sigma_i(j)}\}_{j=1}^{f_t}$ they are correlated with.

Remark If f_t correlated states mutually satisfy $\mathbb{E}(\langle \omega_i | \omega_{f_i(j)} \rangle) < \xi^{-t}$ and $\delta \gg (\xi^t - f_t)^{-1}$, then $\gamma \geq 0$ and so only $\mathcal{O}((\xi^t - f_t)\delta^{-2})$ states are necessary in the sparsification of $|\psi\rangle$ in order that $\mathbb{E}(\||H^{\otimes t}\rangle - |\psi\rangle\|^2) \leq \delta^2$.

Therefore, the scaling with t can be improved by changing the sparsification from i.i.d. sampling to correlated sampling, where each independently sampled state $|\tilde{x}\rangle$ is *supplemented* with f_t states $|\tilde{y}_i\rangle$ that satisfy Remark 1 and thereby produce a sparsfication consisting of only $\mathcal{O}((\xi^t - f_t)\delta^{-2})$ total states. Recently, as a proof of principle, we showed how to accomplish this with $f_t = t$ [5]. To begin, here we extend this linear result to $f_t = 2t - 1$:

Lemma 1 ($2t - 1$ Generation) *Given a t-bitstring, there exist $2t - 1$ additional bitstrings that mutually differ from each other and the given bitstring by at least $t/2$ bitflips.*

The proof of Lemma 1 can be found in Appendix A along with a proof of the subquadratic runtime required for this supplemental bitstring generation and a numerical demonstration of this fact. The code for its implementation is available in Appendix D.

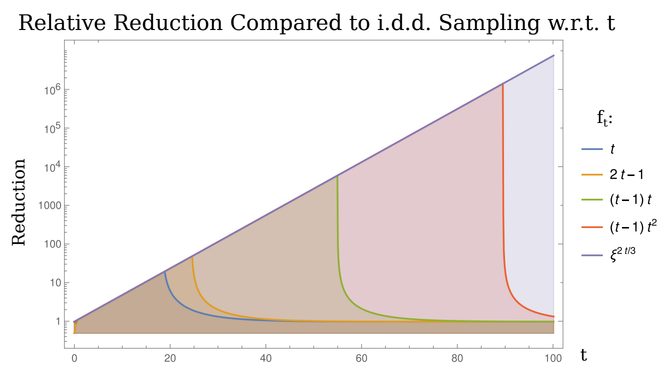

FIG. 2. The relative improvement in scaling w.r.t. t produced by correlated sampling with different $0 \leq f_t \leq \mathcal{O}(\xi^{2t/3})$ compared to the state-of-the-art. The relative improvement is defined as $\max\{\xi^t - f_t, 1\}/\xi^t$, where the maximum is taken since f_t is constrained to be less than ξ_t by Lemma 3 and is integer-valued. In practice, the desired δ determines which f_t you can attain: $\delta^2 \gg (\xi^t - f_t)^{-1} \implies f_t \ll \delta^{-2} - \xi^t$.

It is easy to show that $\eta = 1/2$ is the fraction for which the largest set of bitstrings that are bitflipped with respect to a given bitstring by ηt-bits can be made since this is the value of η that maximizes $\binom{t}{\eta t}$. Note that

Lemma 1 promises a fewer number $2t-1 < \binom{t}{\eta t}$ of such bitstrings because of the additional constraint that the bitstrings *mutually* differ by the same number of bitstrings.

To extend this set and increase f_t properly, it is sufficient to satisfy Remark 1's constraint with added bitstrings, i.e. their *average* inner product with each other must be $\leq \xi^{-t}$, a weaker condition than Lemma 1's constraint of $t/2$ bitflips; the additional bitrsings can mutually differ by more or less than $t/2$ bitflips, though on average they must be fairly close to $t/2$ bitflips.

To accomplish this, we can use the set generated from Lemma 1 as a "seed" set that we generate additional bitstrings from by bitflippling $0 \leq \eta \leq 1/4$ of their bits again, where the threshold η is determined by satisfying Remark 1. This is shown in Lemma 2, and thereby exponentially extends f_t further to $\mathcal{O}(\xi^{2t/3})$:

Lemma 2 *Given a bitstring, there exist $\mathcal{O}(\xi^{2t/3})$ bitstrings where the expectation value of their mutual inner products is $\leq \xi_{\pi/4}^{-t}$ in the limit that $t \to \infty$.*

The proof can be found in Appendix B.

In the construction provided in the proof of Lemma 2, the correlated groups scale as $(2t-1)t^m$. This means that to experience a quadratic benefit in scaling, we will see that the second correlated group can be as large as $2t^2-t$ supplemental states. Ever larger and larger groups of f_t states can be found up until $f_t \in \mathcal{O}(\xi^{2t/3})$. See Figure 2 for a plot of the relative (multiplicative) improvement in scaling this produces w.r.t. t.

How far can you push this technique? While it is trivial to see that $f(t) \in \mathcal{O}(\xi^t)$ since the scaling cost must be non-negative, we proceed to prove a stronger bound of $f(t) \in o(\xi^t)$:

Lemma 3 (Exponential Decay in Supplementation) *The scaling of SPARSIFY is $\mathcal{O}((\xi^t - f_t)\delta^{-2})$ when supplemented with f_t correlated states for each independent state. If #P-hard is greater than BQP-complete, then $f_t \in o(\xi^t)$.*

The proof of Lemma 3 can be found in Appendix C. The result of Lemma 3 means that classical quantum outcome simulation based on the minimal L_1 stabilizer decomposition of the T gate cannot be further improved beyond an additive exponential factor $o(\xi^{t/3})$ with respect to t scaling.

A technical point: as mentioned earlier, there is a conjecture that classical quantum outcome simulation of the T gate magic state with stabilizer states requires $\Omega(\xi^t)$ Gaussian eliminations [1]. A cursory read-through of the results of this paper might lead one to conclude that Lemmas 2 and 3 prove this conjecture since they show that asymptotically classical simulation must scale as $\mathcal{O}(\xi^t)$. However, this is not formally true; the lemmas are constrained to consider a particular classical simulation method, namely one that uses the minimal L_1 stabilizer state decomposition of the T gate magic state given by Eq. 5. Though this is the stabilizer decomposition that minimizes its stabilizer extent, this constraint prevents this result from being a formal proof of the conjecture.

An appealing feature of the algorithm constructively described in Lemmas 1 and 2 is that it can be easily added to the SPARSIFY algorithm used by many contemporary "weak" simulators. It can thus be implemented in current applications with minimal disruption to reduce their sampling cost by an additive exponential constant.

Moreover, a similar approach will also extend this method of correlated L_1 norm sampling to any other diagonal states (see Eq. 1 in [5]). Such a treatment would differ only in that the distribution of t bit-flipped bitstrings would be sampled from the non-uniform distributions.

In conclusion, we constructively prove that classical simulation of quantum outcomes to additive error δ can have a next-order exponential reduction in their sampling cost of $\mathcal{O}((\xi^t - f_t)\delta^2)$ when $\delta^2 \gg (\xi^t - f_t)^{-1}$ and $f_t = \xi^{2t/3}$. We prove that this reduction is nearly asymptotically tight, in that f_t can only be maximally increased by $o(\xi^{t/3})$. This reduction is based on supplementing independent L_1 sampling with correlated sampling of "dissimilar" states and functionally works by further supplementing a "seed" set of $2t-1$ supplemental states. To our knowledge, this work and the work it is based on [5], constitute the first weak simulation algorithm that has lowered this bound's dependence on finite t in the worst-case. It holds great promise in rendering classical simulation of NISQ outcomes more tractable on today's classical computers.

[1] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Physical review letters*, 116(25):250501, 2016.

[2] James R Seddon, Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl T Campbell. Quantifying quantum speedups: improved classical simulation from tighter magic monotones. *arXiv preprint arXiv:2002.06181*, 2020.

[3] Hakop Pashayan, Oliver Reardon-Smith, Kamil Korzekwa, and Stephen D Bartlett. Fast estimation of outcome probabilities for quantum circuits. *arXiv preprint arXiv:2101.12223*, 2021.

[4] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019.

[5] Lucas Kocia. Improved weak simulation of universal quantum circuits by correlated l_1 sampling. *arXiv preprint arXiv:2104.07250*, 2021.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under the Accelerated Research in Quantum Computing program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

This research was supported in part by an appointment to the National Nuclear Security Administration (NNSA) Minority Serving Institutions Internship Program (MSIIP) sponsored by the NNSA and administered by the Oak Ridge Institute for Science and Education (ORISE).

ACKNOWLEDGMENTS

L.K. and G.T. acknowledge helpful discussions with Mohan Sarovar.

Appendix A: Linear Scaling of f_t

A.

t	given bitstring	depth i	t additional bitstrings
2	11	0	α, β
4	1111	0	$\alpha\alpha$, $\beta\beta$
		1	$\alpha\beta$, $\beta\alpha$
8	11111111	0	$\alpha\alpha\alpha\alpha$, $\beta\beta\beta\beta$
		1	$\alpha\alpha\beta\beta$, $\beta\beta\alpha\alpha$
		2	$\alpha\beta\beta\alpha$, $\beta\alpha\alpha\beta$, $\alpha\beta\alpha\beta$, $\beta\alpha\beta\alpha$
16	1111111111111111	0	$\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha$, $\beta\beta\beta\beta\beta\beta\beta\beta$
		1	$\alpha\alpha\alpha\alpha\beta\beta\beta\beta$, $\beta\beta\beta\beta\alpha\alpha\alpha\alpha$
		2	$\alpha\alpha\beta\beta\alpha\alpha\beta\beta$, $\beta\beta\alpha\alpha\beta\beta\alpha\alpha$, $\alpha\alpha\beta\beta\beta\beta\alpha\alpha$, $\beta\beta\alpha\alpha\alpha\alpha\beta\beta$
		3	$\alpha\beta\alpha\beta\alpha\beta\alpha\beta$, $\alpha\beta\alpha\beta\beta\alpha\beta\alpha$, $\beta\alpha\beta\alpha\alpha\beta\alpha\beta$, $\beta\alpha\beta\alpha\beta\alpha\beta\alpha$, $\alpha\beta\beta\alpha\alpha\beta\beta\alpha$, $\beta\alpha\alpha\beta\beta\alpha\alpha\beta$, $\alpha\beta\beta\alpha\beta\alpha\alpha\beta$, $\beta\alpha\alpha\beta\alpha\beta\beta\alpha$

B.

t	given bitstring	depth i	t additional bitstrings
2	11	0	δ
4	1111	0	$\delta\delta$
		1	$\gamma\delta$, $\delta\gamma$
8	11111111	0	$\delta\delta\delta\delta$
		1	$\gamma\gamma\delta\delta$, $\delta\delta\gamma\gamma$
		2	$\gamma\delta\delta\gamma$, $\delta\gamma\gamma\delta$, $\gamma\delta\gamma\delta$, $\delta\gamma\delta\gamma$
16	1111111111111111	0	$\delta\delta\delta\delta\delta\delta\delta\delta$
		1	$\gamma\gamma\gamma\gamma\delta\delta\delta\delta$, $\delta\delta\delta\delta\gamma\gamma\gamma\gamma$
		2	$\gamma\gamma\delta\delta\gamma\gamma\delta\delta$, $\delta\delta\gamma\gamma\delta\delta\gamma\gamma$, $\gamma\gamma\delta\delta\delta\delta\gamma\gamma$, $\delta\delta\gamma\gamma\gamma\gamma\delta\delta$
		3	$\gamma\delta\gamma\delta\gamma\delta\gamma\delta$, $\gamma\delta\gamma\delta\delta\gamma\delta\gamma$, $\delta\gamma\delta\gamma\gamma\delta\gamma\delta$, $\delta\gamma\delta\gamma\delta\gamma\delta\gamma$, $\gamma\delta\delta\gamma\gamma\delta\delta\gamma$, $\delta\gamma\gamma\delta\delta\gamma\gamma\delta$, $\gamma\delta\delta\gamma\delta\gamma\gamma\delta$, $\delta\gamma\gamma\delta\gamma\delta\delta\gamma$

TABLE I. $\alpha \equiv 01$, $\beta \equiv 10$, $\gamma \equiv 11$, and $\delta \equiv 00$. The additional bitstrings can be used as XOR masks to generate the appropriate additional bitstrings for given bitstrings other than all 1s.

An algorithm that generates $2t-1$ additional bitstrings that differ by at least $t/2$ bitflips, such as those given in Table 1, is given in Algorithm 1. A. is the previous algorithm that generates t additional bitstrings, and B. is the addition to this algorithm that generates $t-1$ more.

Algorithm 1: Generate additional bitstrings that differ from the t-bitstring of all 1s by at least $t/2$ bitflips.

Data: k such that $t = 2^k$.
Result: *bitstring* array.
begin
  bitstrings $\leftarrow$ A. $\{\alpha \cdots \alpha,\ \beta \cdots \beta\}$ or B. $\{\delta \cdots \delta\}$;
  masks $\leftarrow \{2^{2^k}\}$;
  for *treedepth* $\leftarrow 1$ **to** $k-1$ **do**
    levelmask $\leftarrow 2^{2^{k-treedepth}} - 1$;
    for *levelmaskdepth* $\leftarrow 2$ **to** $2^{treedepth-1}$ **do**
      levelmask $\leftarrow$ *levelmask* $+\ 2^{2^{k-treedepth+1}} \times$ *levelmask*;
    end
    A. *bitstring* $\leftarrow (\alpha \cdots \alpha\ \mathrm{XOR}\ levelmask)$ or
    B. *bitstring* $\leftarrow (\gamma \cdots \gamma\ \mathrm{XOR}\ levelmask)$;
    for *mask* $\leftarrow$ *Cartesian products of elements in masks* **do**
      add (*bitstring* XOR *mask*) to *bitstrings*;
    end
    add *levelmask* to *masks*;
  end
end

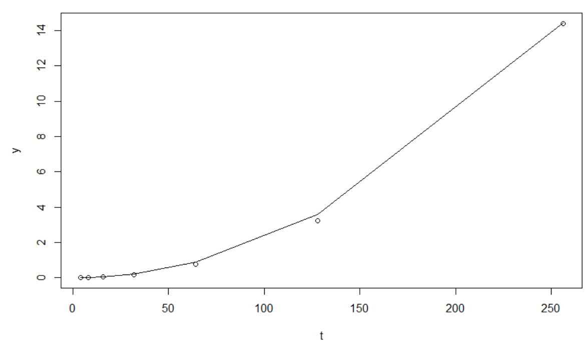

FIG. 3. Plot demonstrating quadratic runtime scaling of Algorithm 1 given by R code in Appendix. (dots) raw runtime, (line) quadratic fit. Technically the runtime is $\mathcal{O}(\log(t)t^2)$ since every point plotted above depends on running the previous point (i.e. this is a differential plot and should be a cumulative plot). Since there are $\log_2(t)$ points that go into every point with less than t^2 runtime, the full scaling is multiplied by $\log(t)$. Finally, since this runtime generates $2t-1$ bitstrings, the overall scaling should be divided by $\mathcal{O}(t)$. This means that the overall runtime is $\mathcal{O}(\log(t)t)$ to generate each additional bitstring.

Lemma 1 **($2t-1$ Generation)** *Given a t-bitstring, there exist $2t-1$ additional bitstrings that mutually differ from each other and the given bitstring by at least $t/2$ bitflips.*

Proof Let $\alpha = \{10\}$, $\beta = \{01\}$, $\gamma = \{00\}$, $\delta = \{11\}$. WLOG we consider x as the bitstring of all 1's since we can always consider supplementing any other given bitstring by XORing the generated additional bitstrings.

Given $x \cdot y_i, y_i \cdot y_j \leq \frac{1}{\sqrt{2}^{\frac{2^k}{2}}}$ for $i \neq j$ and $t = 2^k$. When $t = 2$ the y's are α, β and γ. These satisfy the previous inequalities.

Assume true for $t = 2^k$, for some $k > 1$. $\frac{2^k}{2} = 2^{k-1}$; $\frac{2^{k+1}}{2} = 2^k$ For a bitstring length of 2^k, we assume we generate $(2(2^k) - 1)$ y_i's $\rightarrow (2^{k+1} - 1)$ y_i's.

We now consider $t = 2^{k+1}$. For a bitstring length of 2^{k+1}, we want $(2(2(2^k)) - 1)$ y_i's $\rightarrow 2(2^{k+1}) - 1)$ y_i' s. We define the operation cloning. Cloning x allow us to get xx where $x \rightarrow xx$ taking our bitstring length $2^k \rightarrow 2(2^k) = 2^{k+1}$ and similarly, cloning y_i allow us to get $y_i y_i$ where $y_i \rightarrow y_i y_i$ taking our bitstring length $2^k \rightarrow 2(2^k) = 2^{k+1}$.

After cloning our new equation is now

$$xx \cdot y_i y_i, y_i y_i \cdot y_j y_j \leq \frac{1}{\sqrt{2}^{\frac{2^{k+1}}{2}}}, i \neq j \tag{A1}$$

Our m-rule states that $a \cdot b$ means how many bits are different between a and b, which is our inner product multiplicative.

$$m = m_1 + m_2; \frac{1}{\sqrt{2}^{m_1+m_2}} = \frac{1}{\sqrt{2}^{m_1}} * \frac{1}{\sqrt{2}^{m_2}} \tag{A2}$$

Complementation takes $y_i \rightarrow y_j$ where $j \neq i$. Define $\overline{\alpha} = \beta$, $\overline{\beta} = \alpha$, $\overline{\gamma} = \delta$, and $\overline{\delta} = \gamma$.

We now consider the t additional bitstrings produced from all cloned bitstrings, which additionally have none or one of their components complemented. From our assumption, there are $2(2^k) - 1$ bitstrings produced with no complemented components. Complementation of the lower bits produces $2(2^k) - 1$ additional bitstrings. This brings us to a total of $2(2(2^k) - 1) = 2(2^{k+1}) - 2$. The final additional bitstring we consider is of all γ's bringing the total to $2(2^{k+1}) - 1 = 2t - 1$ additional prospective bitstrings. We now proceed to show that these are all valid.

$xx \equiv x \cdot 2^{2^k} + x$. Similarly, $y_i y_i \equiv y_i \cdot 2^{2^k} + y_i$. $xx \cdot y_i y_i \rightarrow (x \cdot 2^{2^k} + x) \cdot (y_i \cdot 2^{2^k} + y_i) = (x \cdot 2^{2^k}) \cdot (y_i \cdot 2^{2^k}) * (x \cdot y_i)$. We know that $(x \cdot 2^{2^k}) \cdot (y_i \cdot 2^{2^k}) \leq \frac{1}{\sqrt{2}^{2^{k-1}}}$ and $(x \cdot y_i) \leq \frac{1}{\sqrt{2}^{2^{k-1}}}$. Therefore $\frac{1}{\sqrt{2}^{2^{k-1}}} * \frac{1}{\sqrt{2}^{2^{k-1}}} = \frac{1}{\sqrt{2}^{2^k}}$

$xx \equiv x \cdot 2^{2^k} + x$. Similarly, $y_i y_j \equiv y_i \cdot 2^{2^k} + y_j$. $xx \cdot y_i y_j \rightarrow (x \cdot 2^{2^k} + x) \cdot (y_i \cdot 2^{2^k} + y_j) = (x \cdot 2^{2^k}) \cdot (y_i \cdot 2^{2^k}) * (x \cdot y_j)$. We know that $(x \cdot 2^{2^k}) \cdot (y_i \cdot 2^{2^k}) \leq \frac{1}{\sqrt{2}^{2^{k-1}}}$ and $(x \cdot y_j) \leq \frac{1}{\sqrt{2}^{2^{k-1}}}$. Therefore $\frac{1}{\sqrt{2}^{2^{k-1}}} * \frac{1}{\sqrt{2}^{2^{k-1}}} = \frac{1}{\sqrt{2}^{2^k}}$

$y_i y_i \equiv y_i \cdot 2^{2^k} + y_i$. Similarly, $y_j y_j \equiv y_j \cdot 2^{2^k} + y_j$. $y_i y_i \cdot y_j y_j \rightarrow (y_i \cdot 2^{2^k} + y_i) \cdot (y_j \cdot 2^{2^k} + y_j) = (y_i \cdot 2^{2^k}) \cdot (y_j \cdot 2^{2^k}) * (y_i \cdot y_j)$. We know that $(y_i] \cdot 2^{2^k}) \cdot (y_j \cdot 2^{2^k}) \leq \frac{1}{\sqrt{2}^{2^{k-1}}} and (y_i \cdot y_j) \leq \frac{1}{\sqrt{2}^{2^{k-1}}}$. Therefore $\frac{1}{\sqrt{2}^{2^{k-1}}} * \frac{1}{\sqrt{2}^{2^{k-1}}} = \frac{1}{\sqrt{2}^{2k}}$.

Appendix B: Exponential Scaling of f_t

Lemma 2 *Given a bitstring, there exist $\mathcal{O}(\xi^{2t/3})$ bitstrings where the expectation value of their mutual inner products is $\leq \xi_{\pi/4}^{-t}$ in the limit that $t \rightarrow \infty$.*

Proof Take the $2t - 1$ bitstrings from Lemma 1, and the given bitstring, and bitflip $M < t/4$ bits. This can be done to any of the t bits for each bitstring. The resulting $\binom{t}{M}$ bitstrings differ from each other by τ bitflips, where $t/2 - 2M \leq \tau \leq t/2 + 2M$. For every bitstring that differs from one by $t/2 - \alpha$, there exists another that differs from it by $t/2 + \alpha$.

It follows that

$$\begin{aligned} \mathbb{E}(\langle \omega_i | \omega_{f_i(j)} \rangle &\geq \frac{1}{2} \left(2^{-t/4 - 2(M-1)/2} + 2^{-t/4 + 2(M-1)/2} \right) \\ &= 2^{-t/4 + M - 2}(1 + 2^{-2M})). \end{aligned} \tag{B1}$$

In the limit that $t \rightarrow \infty$, this is less than $\xi_{\pi/4}^{-t} = \cos(\pi/8)^{-t/2} = 2^{2 \log_2 \cos(\pi/8) t}$ if $M < (2 \log_2 \cos(\pi/8) + 1/4)t \lesssim 0.02t$.

It follows that

$$\begin{aligned} f_t \quad &< \quad \binom{t}{M} 2t \\ &\lim_{t \rightarrow \infty} \propto \sqrt{t} 2^{\sim 0.15t} \in \mathcal{O}(\xi^{2t/3}). \end{aligned} \tag{B2}$$

■

The bound on the supplemental bitstrings in Lemma 2 is meant to be easy to prove, and is in fact far more favorable in practice. Our aim was to find an exponential improvement in f_t not to maximize its exponential factor.

Lemma 2 is constructive and proves that each of the $f_t - (2t - 1)$ additional bitstrings can be generated by supplementing the bitstrings generated by Lemma 1 in $\mathcal{O}(f_t - (2t - 1))$ time.

Appendix C: Asymptotic Limit

Lemma 3 (Exponential Decay in Supplementation) *The scaling of SPARSIFY is $\mathcal{O}((\xi^t - f_t)\delta^{-2})$ when supplemented with f_t correlated states for each independent state. If #P-hard is greater than BQP-complete, then $f_t \in o(\xi^t)$.*

Proof If #P-hard is greater than BQP-complete then,

$$\chi^{2t} \underset{t \rightarrow \infty}{\geq} (\xi^t - f_t)\delta^{-2}, \tag{C1}$$

for $\delta^2 \gg (\xi^t - f_t)^{-1}$ and where χ^t is the stabilizer rank of the T gate magic state.

$$\begin{aligned} \implies \quad f_t &\geq \xi^t - \chi^{2t}\delta^2 \\ f_t &\gg \xi^t - \chi^{2t}(\xi^t - f_t)^{-1}, \end{aligned} \tag{C2}$$

$$\Leftrightarrow f^2(t) \gg \chi^{2t} - \xi^{2t}. \tag{C3}$$

This implies that

$$\lim_{t\to 0} \frac{f^2(t)}{\xi^{2t}} \gg \lim_{t\to 0} \frac{\chi^{2t}}{\xi^{2t}} - 1 = 0. \tag{C4}$$

Therefore,

$$\lim_{t\to 0} \frac{f_t}{\xi^t} = 0, \tag{C5}$$

and so $f_t \in o(\xi^t)$. ■

Appendix D: R Code for $2t - 1$ Supplemental State Generation

```
alpha <- c(1,0)
beta <-c(0,1)
gamma <-c(0,0)
delta <-c(1,1)

cloning <- function(vec)
{
  return <- array(c(vec,vec))
}
comp <-function(vec)
{
  for(i in 1:length(vec))
  {
    vec[i]= (vec[i]+ 1) %% 2
  }
  return <- vec
}
comp2 <-function(vec)
{
  for(i in (length(vec)/2 +1):length(vec))
  {
    vec[i]= (vec[i]+ 1) %% 2
  }
  return <- vec
}

dp <- function(vecx, vecy)
{
  m <- 0
  for (i in 1:length(vecx))
  {
  if (vecx[[i]] != vecy[[i]]) m <- m+1
  }

  val <- (1/sqrt(2)^(m))
  return <- val
}

yis <-list(alpha, beta, gamma)

for(k in 2:8)
{
  print(k)
  oldyis<-yis

  for (l in 1:100)
  {
    yis<- oldyis
  x<-c()
  for (j in 1:(2^k))
  {
    x<- array(c(1,x))
  }

  newyis <-list()

  for(i in 1:length(yis))
  {
    newyis[[i]]<-(cloning(yis[[i]]))
  }
  for(i in 1:length(yis))
  {
    newyis[[i+length(yis)]]<-(comp2(cloning(yis[[i]])))
  }

  newyis[[2*length(yis)+1]]<-comp(newyis[[2*length(yis)]]
  yis<- newyis
  }
}
```

Technical Paper #2 of 2

Improved Weak Simulation of Universal Quantum Circuits by Correlated L_1 Sampling

Lucas Kocia
Sandia National Laboratories, Livermore, California 94550, U.S.A.

Bounding the cost of classically simulating the outcomes of universal quantum circuits to additive error δ is often called weak simulation and is a direct way to determine when they confer a quantum advantage. Weak simulation of the T+Clifford gateset is BQP-complete and is expected to scale exponentially with the number t of T gates. We constructively tighten the upper bound on the worst-case L_1 norm sampling cost to next order in t from $\mathcal{O}(\xi^t \delta^{-2})$ if $\delta^2 \gg \xi^{-t}$ to $\mathcal{O}((\xi^t - t)\delta^{-2})$ if $\delta^2 \gg (\xi^t - t)^{-1}$, where $\xi^t = 2^{\sim 0.228t}$ is the stabilizer extent of the t-tensored T gate magic state. We accomplish this by replacing independent L_1 sampling in the popular SPARSIFY algorithm used in many weak simulators with correlated L_1 sampling. As an aside, this result demonstrates that the T gate magic state's approximate stabilizer state decomposition is not multiplicative with respect to t, for finite values, despite the multiplicativity of its stabilizer extent. This is the first weak simulation algorithm that has lowered this bound's dependence on finite t in the worst-case to our knowledge and establishes how to obtain further such reductions in t.

Weak simulation is defined as the task of sampling the probabilities of universal quantum circuits to additive error. It is expected to require exponential resources on a classical computer since it is BQP-complete. Reducing the cost of classically simulating quantum computers [1] is necessary to characterize near-term noisy intermediate-scale quantum (NISQ) computers [2] that are rapidly growing in size and performance.

Universal quantum computation can be achieved using stabilizer states, the Clifford+T gateset, and Pauli measurement. An equivalent measurement-based formalism can be written in terms of stabilizer states, T gate magic states and Pauli measurements [3]. Approximating outcomes samples to additive error δ is equivalent to replacing the underlying probability distribution with one that is δ-close to it and then sampling from this approximate distribution. This naturally splits up many weak simulation implementations into a "sparsification" step and a measurement step. The measurement step consists of taking idempotent projections, $\langle\psi|\Pi|\psi\rangle = |\Pi|\psi\rangle|^2 \equiv |\psi'|^2$, and so is frequently called a "normalization" step instead.

The SPARSIFY algorithm introduced [4] a method of generating an L_1 sparsification of a given state ψ to δ additive error with $\mathcal{O}(2^{\sim 0.228t}\delta^{-2})$ stabilizer states, which is asymptotically optimal as $t \to \infty$ and $\delta \to 0$ (see Lemma 2 in [4]). As a result, the authors conjectured the following lower bound:

Conjecture 1 *Any approximate stabilizer decomposition of $T^{\otimes t}$ that achieves a constant approximation error must use at least $\Omega(2^{\sim 0.228t})$ stabilizer states.*

This approximated state's inner product must then be sampled under random Pauli measurements to complete a weak simulation algorithm. The full weak simulation cost is the number of stabilizer states produced by SPARSIFY multiplied by $\mathcal{O}(t^3\delta^{-2})$.

Subsequent works [5–7] almost all use the SPARSIFY algorithm or a similar sparsification method. Improvements have included an extension of the method to diagonal states (other than the T magic state) [5] and mixed states [6], constant factor improvements [5], a decrease in the power of δ cost for magic states [6], better performance when the values of the sampled probabilities are in certain regimes [7], and an extension to Born probabilities [7].

Nevertheless, these methods have all saturated the asymptotic conjectured lower bound w.r.t. t even when they are not in the asymptotic limit of $t \to \infty$ and $\delta \to 0$; they all require $\mathcal{O}(2^{\sim 0.228t})$ stabilizer states in the worst case.

Since the lower bound given by Conjecture 1 is an asymptotic bound, there is no reason to consider it limiting for finite t and δ. Indeed, the finite regime is the most useful for practical simulations and validations of near-term devices. Non-asymptotic reductions in t can greatly increase the size of universal quantum circuits that are simulatable by today's classical computers and thereby change when they confer quantum advantage.

Here we introduce the first such reduction in t and demonstrate its practical usefulness for finite-sized circuits. Since the reduction occurs in the SPARSIFY algorithm used by many contemporary weak simulators, it can be implemented in current applications with minimal change and improve their performance. The key idea is replacing independent L_1 sampling with correlated L_1 sampling.

To begin, we define the general family of diagonal states that we want to approximate. Following [3, 4], we define the t-tensored state:

$$|D_\phi^{\otimes t}\rangle \equiv (2\nu)^{-t} \sum_{x\in\mathbb{F}_2^t} |\tilde{x}_1 \otimes \cdots \otimes \tilde{x}_t\rangle, \tag{1}$$

where

$$|\tilde{0}\rangle \equiv \frac{i}{\sqrt{2}}(-i + e^{-\pi i/4})(-i + e^{i\phi})|0\rangle, \tag{2}$$

$$|\tilde{1}\rangle \equiv \frac{i}{\sqrt{2}}(1 + e^{-\pi i/4})(1 - e^{i\phi})\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \tag{3}$$

and $\nu \equiv \cos\pi/8$. $|D_\phi^{\otimes t}\rangle$ has L_1 norm squared, when considered over the set of stabilizer states, that is minimized by this stabilizer decomposition into $|\tilde{0}\rangle$ and $|\tilde{1}\rangle$:

$$\left(\sum_{x\in\mathbb{F}_2^t} |c_{\tilde{x}}|\right)^2 = \left(\sqrt{1-\sin\phi}+\sqrt{1-\cos\phi}\right)^{2t} \equiv \xi_\phi^t. \quad (4)$$

Calling ξ_ϕ^t an L_1 norm is an abuse of the term, since stabilizer states are overcomplete. As a result ξ_ϕ^t is often called the *stabilizer extent*, and is more accurately defined as the minimum value of $\|c\|_1$ over all stabilizer decompositions.

$|D_{\pi/4}^{\otimes t}\rangle = |H^{\otimes t}\rangle$ for $|H\rangle \equiv e^{-i\pi/8}SH|T\rangle$ and

$$|T\rangle = \frac{1}{\sqrt{2}}(|0\rangle + \sqrt{i}|1\rangle), \quad (5)$$

where S and H are the Clifford gates phase shift and Hadamard, respectively. $\xi_{\pi/4}^t = \nu^{-2t} = 2^{\sim 0.228t}$, which sets the exponential scaling for sampling this magic state. Since the H magic state is related to the T magic state by a Clifford unitary under which stabilizer states are closed, its stabilizer state approximation cost is the same.

Before we get into generating an approximation to these states, we first need to establish a useful result:

Lemma 1 (t Stabilizer States with $t/2$ Bit Flips) *Given a t-bitstring, there exist t additional different t-bitstrings such that every pair of bitstrings differs by at least $t/2$ bits.*

Proof We examine powers of 2. WLOG we can always assume that the given bitstring consists of all 1s since any additional bitstrings to this can be generalized to any other bitstring by XORing it.

Let $t = 2^k$. Consider iterative splitting of the 2^k bitstring by binary tree with layers $0 \le i \le k-1$ (see Table 1 for examples). We define the base layer $k = 0$ to consist of two bitstrings, $\alpha\cdots\alpha$ and $\beta\cdots\beta$, where $\alpha \equiv 01$ and $\beta \equiv 10$. Given a t-bitstring of all 1s, it is clear that these two bitstrings differ from it by $t/2$ bitflips and from each other by $t > t/2$ bitflips.

In every subsequent layer we consider t-bitstrings that are evenly split into $2 \le 2^i \le 2^{k-1}$ contiguous blocks of bits that we treat all together when assigning values. We assign values α and β to the blocks *such that* the number of αs and βs are even in every pair of blocks corresponding to a larger block a layer above and at least half the assignments differ from every other bitstring in the same layer.

This means that every layer $1 \le i \le k-1$ will have $t/2$ αs and $t/2$ βs. Hence, they will differ from the given bitstring of all 1s by $t/2$ bit flips. This also means that they will differ from bitstrings in the zeroth layer consisting of only αs or βs by $t/2$ bit flips.

Moreover, since the nature of this binary tree splitting converts blocks from higher levels into evenly split subblocks with the same number of αs and βs where before there were only αs or βs, every bitstring differs from those of other layers by $t/2$ bitflips.

t	given bitstring	depth i	t additional bitstrings
2	11	0	α, β
4	1111	0	$\alpha\alpha$, $\beta\beta$
		1	$\alpha\beta$, $\beta\alpha$
8	11111111	0	$\alpha\alpha\alpha\alpha$, $\beta\beta\beta\beta$
		1	$\alpha\alpha\beta\beta$, $\beta\beta\alpha\alpha$
		2	$\alpha\beta\beta\alpha$, $\beta\alpha\alpha\beta$, $\alpha\beta\alpha\beta$, $\beta\alpha\beta\alpha$
16	1111111111111111	0	$\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha$, $\beta\beta\beta\beta\beta\beta\beta\beta$
		1	$\alpha\alpha\alpha\alpha\beta\beta\beta\beta$, $\beta\beta\beta\beta\alpha\alpha\alpha\alpha$
		2	$\alpha\alpha\beta\beta\alpha\alpha\beta\beta$, $\beta\beta\alpha\alpha\beta\beta\alpha\alpha$, $\alpha\alpha\beta\beta\beta\beta\alpha\alpha$, $\beta\beta\alpha\alpha\alpha\alpha\beta\beta$
		3	$\alpha\beta\alpha\beta\alpha\beta\alpha\beta$, $\alpha\beta\alpha\beta\beta\alpha\beta\alpha$, $\beta\alpha\beta\alpha\alpha\beta\alpha\beta$, $\beta\alpha\beta\alpha\beta\alpha\beta\alpha$, $\alpha\beta\beta\alpha\alpha\beta\beta\alpha$, $\beta\alpha\alpha\beta\beta\alpha\alpha\beta$, $\alpha\beta\beta\alpha\beta\alpha\alpha\beta$, $\beta\alpha\alpha\beta\alpha\beta\beta\alpha$

TABLE I. $\alpha \equiv 01$ and $\beta \equiv 10$. The additional bitstrings can be used as XOR masks to generate the appropriate additional bitstrings for given bitstrings other than all 1s.

Within a layer, by construction, the bitstrings differ from each other by at least $t/2$ bitflips. Trivially, the first layer ($i = 1$) consists of two bitstrings. Subsequent layers consist of twice as many bitstrings as their preceding layers since they can be considered as the result of the same bitflips performed on the given bitstring as the layer above, but over twice as many bits. It follows that the ith level consists of 2^i bitstrings. This means that there is a total of $2 + \sum_{i=1}^{k-1} 2^i = 2^k = t$ bitstrings.

We have therefore t additional total of bitstrings that differ from the given bitstrings of all 1s and each other by at least $t/2$ bitflips. ∎

An algorithm that generates t additional bitstrings that differ by at least $t/2$ bitflips, such as those given in Table 1, is given in Algorithm 1.

We introduce a constructive upper bound that is lower than Conjecture 1's in the finite t case:

Theorem 1 (Lower Bound in t for SPARSIFY) *The SPARSIFY procedure introduced by Bravyi* et al. *[5] creates a δ-approximate stabilizer decomposition of $H^{\otimes t}$ with $\mathcal{O}((2^{\sim 0.228t}-t)\delta^{-2})$ stabilizer states for t sufficiently large such that $\delta^2 \gg (\xi_{\pi/4}^t - t)^{-1}$.*

Proof Following [5], we define additive error

$$\||D_\phi^{\otimes t}\rangle - |\psi\rangle\| \le \delta, \quad (6)$$

where $\|\psi\| \equiv \sqrt{\langle\psi|\psi\rangle}$.

$|\psi\rangle$ is the sparsified k-term approximation to $|D_\phi(t)\rangle$ given by

$$|\psi\rangle = \frac{\|c\|_1}{k}\sum_{i=1}^{k}|\omega_i\rangle, \quad (7)$$

where each $|\omega_i\rangle$ is independently chosen randomly so that it is a normalized stabilizer state $|\omega_i\rangle = c_i/|c_i||\varphi_i\rangle$ with

Algorithm 1: Generate additional bitstrings that differ from the t-bitstring of all 1s by at least $t/2$ bitflips.

```
Data: k such that t = 2^k.
Result: bitstring array.
begin
    bitstrings ← {α···α, β···β};
    masks ← {2^(2^k)};
    for treedepth ← 1 to k − 1 do
        levelmask ← 2^(2^(k−treedepth)) − 1;
        for levelmaskdepth ← 2 to 2^(treedepth−1) do
            levelmask ← levelmask +
              2^(2^(k−treedepth+1)) × levelmask;
        end
        bitstring ← (α···α XOR levelmask);
        for mask ← Cartesian products of elements in
          masks do
            add (bitstring XOR mask) to bitstrings;
        end
        add levelmask to masks;
    end
end
```

probability $p_i = |c_i|/\|c\|_1$. We define a random variable $|\omega\rangle$ that is equal to $|\omega_i\rangle$ with probability p_i. Then

$$\mathbb{E}(|\omega\rangle) = |\psi\rangle/\|c\|_1. \tag{8}$$

By construction,

$$\mathbb{E}(\langle\psi|D_\phi^{\otimes t}\rangle) = \mathbb{E}(\langle D_\phi^{\otimes t}|\psi\rangle) = 1. \tag{9}$$

The number of stabilizer states in the approximation is k and

$$\begin{aligned}\mathbb{E}(\||D_\phi^{\otimes t}\rangle - |\psi\rangle\|^2) = \mathbb{E}(|\langle D_\phi^{\otimes t}|D^{\otimes t}\rangle|) - \mathbb{E}(|\langle D_\phi^{\otimes t}|\psi\rangle|) - \\ \mathbb{E}(|\langle\psi|D_\phi^{\otimes t}\rangle|) + \mathbb{E}(|\langle\psi|\psi\rangle|) \\ \leq \frac{\xi_\phi^t}{k} - \frac{\gamma}{k},\end{aligned} \tag{10}$$

where we simplified

$$\begin{aligned}\mathbb{E}(\langle\psi|\psi\rangle) &= \sum_i^k \frac{\|c\|_1^2}{k^2}\mathbb{E}(\langle\omega_i|\omega_i\rangle) + \sum_{i\neq j}^k \frac{\|c\|_1^2}{k^2}\mathbb{E}(\langle\omega_i|\omega_j\rangle) \\ &= \frac{\|c\|_1^2}{k}\mathbb{E}(|\langle\omega|\omega\rangle|) + \sum_{i\neq j}^k \frac{\|c\|_1^2}{k^2}\mathbb{E}(\langle\omega_i|\omega_j\rangle) \\ &\leq \frac{\|c\|_1^2}{k} + 1 - \frac{\gamma}{k}.\end{aligned} \tag{11}$$

Eq. 10 is less than or equal to δ^2 when $k = (\xi_\phi^t - \gamma)\delta^{-2}$. If $|\omega_i\rangle$ are independent and identically distributed (i.i.d.) stabilizer states then $\sum_{i\neq j}^k \|c\|_1^2\mathbb{E}(\|\langle\omega_i|\omega_j\rangle) = \sum_{i\neq j}^k |\mathbb{E}(\langle\psi|)\mathbb{E}(|\psi\rangle)| = k(k-1)$ and so $\gamma = 1$. As a result, since $1 \ll \xi_\phi^t$ as t increases, it was neglected in previous characterizations [4].

However, γ can become significant if $|\omega_i\rangle$ are not i.i.d. and $\sum_{i\neq j}^k \|c\|_1^2\mathbb{E}(\|\langle\omega_i|\omega_j\rangle) \neq \sum_{i\neq j}^k |\mathbb{E}(\langle\psi|)\mathbb{E}(|\psi\rangle)|$.

In particular, let us consider sampling the H magic state $|D_{\pi/4}^{\otimes t}\rangle = |H^{\otimes t}\rangle$. In this case, its minimal L_1 stabilizer state decomposition consists of a uniform superposition over $|\tilde{0}\rangle$ and $|\tilde{1}\rangle$, where $|\langle\tilde{0}|\tilde{1}\rangle| = 2^{-\frac{1}{2}}$. In the SPARSIFY algorithm, t-bit strings consisting of these stabilizer states are uniformly sampled to approximate $|H^{\otimes t}\rangle$.

By Lemma 1, let us supplement this set of t-bit strings with the t t-bit strings that differ from every uniformly sampled state and each other by at least $t/2$ bit flips (referring to the tilde basis). It follows that these $(t+1)$ stabilizer states have inner products of $\leq 2^{-\frac{t/2}{2}}$.

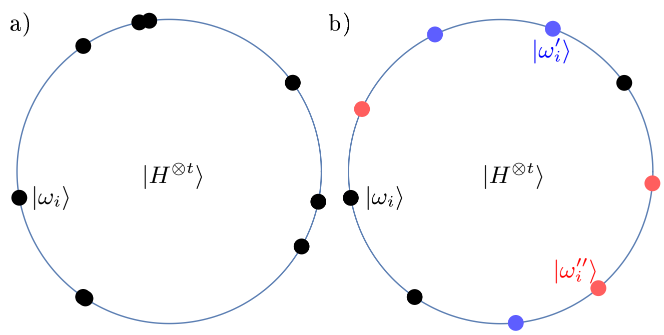

FIG. 1. Sketch of the different ensembles produced by (a) independent and (b) correlated L_1 sampling. In this example, (a) there are nine independently sampled states, $\{|\omega_i\rangle\}_i$, from the uniform distribution on the unit circle (the "stabilizer state space"), which when considered as real vectors on $\mathbb{R}^2$ have an expectation value close to $|H^{\otimes t}\rangle$ at the origin. This ensemble can be transformed to a (b) correlated one, by supplementing the first three states with two $\frac{t}{2}$-bitflipped versions, $|\omega_i'\rangle$ and $|\omega_i''\rangle$ (blue and red), which are therefore far away and equidistant to each other on the unit circle, and discarding the rest. The expectation value of this ensemble is closer to $|H^{\otimes t}\rangle$ but converges more slowly.

Since the ensemble consisting of uniformly sampled t-bit strings $|\omega_i\rangle$ satisfies

$$\mathbb{E}(\langle\psi|H^{\otimes t}\rangle) = \mathbb{E}(\langle H^{\otimes t}|\psi\rangle) = 1, \tag{12}$$

it follows that the t other ensembles consisting of the ith state with at least $t/2$ bits flipped ($i \in \{1, \ldots, t\}$) compared to the uniformly sampled states, also satisfy this property. Therefore, the full ensemble produced by adding together these $(t+1)$ ensembles satisfies this property too.

However, taken together, these are no longer i.i.d. stabilizer states. In particular, for a given $\langle\omega_i|$, there exist at least t $|\omega_{f_i(j)}\rangle$ such that $|\langle\omega_i|\omega_{f_i(j)}\rangle| \le 2^{-\frac{t/2}{2}}$. Hence,

$$\begin{aligned}
\sum_{i\neq j}^{k} \|c\|_1^2 \mathbb{E}(\langle\omega_i|\omega_j\rangle) &= \sum_{i}^{k} \sum_{\substack{j \\ i\neq j}}^{k-t} \|c\|_1^2 \mathbb{E}(\langle\omega_i|)\mathbb{E}(|\omega_j\rangle) \quad (13)\\
&+ \sum_{i}^{k}\sum_{j}^{t} \|c\|_1^2 \mathbb{E}(\langle\omega_i|\omega_{f_i(j)}\rangle)\\
&\le k(k-1-t) + k\|c\|_1^2 2^{-\frac{t/2}{2}} t \quad (14)\\
&= k(k-\gamma), \quad (15)
\end{aligned}$$

where $\|c\|_1^2 = \xi_{\pi/4}^t = 2^{\sim 0.228t}$ and so $\gamma = 1 + (1 - 1/2^{\sim 0.02t})t$.

Therefore, given that at least $k = (\xi_\phi^t - \gamma)\delta^{-2}$ stabilizer states are necessary to sample this state to δ additive error, the SPARSIFY procedure creates a δ-approximate stabilizer decomposition of $H^{\otimes t}$ with $\mathcal{O}((2^{\sim 0.228t} - 1 - (1 - 1/2^{\sim 0.022t})t)\delta^{-2})$ stabilizer states.

This more efficiently approximated state comes at the expense of its convergence probability, or sparsification tail bound. Following the same reasoning as in the proof of Lemma 7 of [5],

$$\begin{aligned}
&\Pr\left[\|H^{\otimes t} - \psi\|^2 \le \langle\psi|\psi\rangle - 1 + \delta^2\right]\\
&\ge 1 - 2\exp\left(-\frac{\delta^2\xi_{\pi/4}^t}{8} + \frac{\gamma\delta^2}{8}\right) \quad (16)\\
&= 1 - 2\exp\left(-\frac{\delta^2\xi_{\pi/4}^t}{8} + \frac{(1+(1-1/2^{\sim 0.02t})t)\delta^2}{8}\right). \quad (17)
\end{aligned}$$

Therefore, given that $\delta^2 \gg (\xi_{\pi/4}^t - t)^{-1}$, if post-selection is performed to discard samples that produce $\langle\psi|\psi\rangle - 1 \gg \delta^2$ (a rare event if this first condition is met) or $\langle\psi|\psi\rangle$ is approximated to relative error using the FASTNORM algorithm [5] (which scales linearly with k), then the states ψ are generated with $\mathbb{E}(\||H^{\otimes t}\rangle - |\psi\rangle\|^2) < \delta^2$ and consist of $\mathcal{O}((2^{\sim 0.228t} - t)\delta^{-2})$ stabilizer states. ■

A sketch of the key idea used in the proof of Theorem 1 is shown in Figure 1. Independently sampled states with expectation value $|H^{\otimes t}\rangle$ are replaced with a smaller subset that are supplemented with bit-flipped states. The resultant correlated distribution has an expectation value closer to $|H^{\otimes t}\rangle$, but it converges to it more slowly.

Some polynomial factors in t are not included in the scaling cost, $\mathcal{O}((2^{\sim 0.228t} - t)\delta^{-2})$, of SPARSIFY. Moreover, there is a possible additional polynomial cost in correlated sampling from generating the bit-flipped supplemental states (such as using Algorithm 1) compared to independent sampling. We claim that these changes in polynomial factors are negligible. This claim is supported by the scaling observed in the practical runtime of SPARSIFY plotted in Figure 2. A decrease in runtime is observed for correlated sampling that is lower bounded by proportionality to the fewer number of stabilizer states $k = (\xi_\phi^t - \gamma)\delta^{-2}$ it generates.

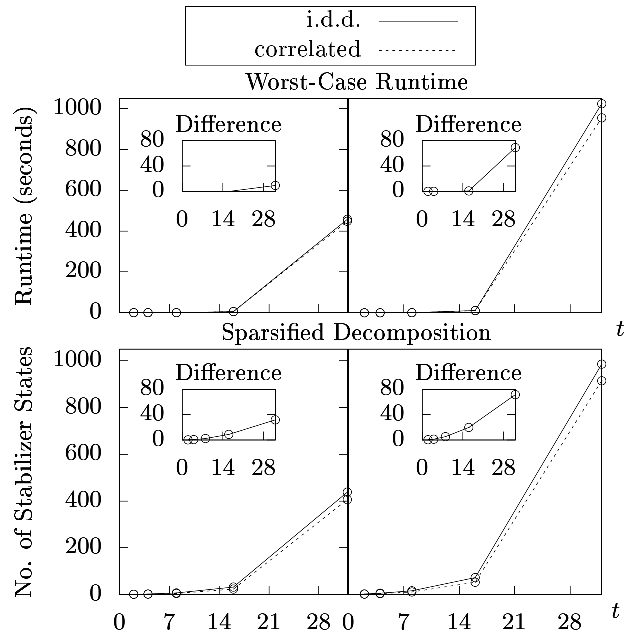

FIG. 2. Plots are for additive error (left) $\delta = 0.6$ and (right) $\delta = 0.4$ over 100 runs. The worst-case runtime of calculating the norm of ψ where $|\psi - T^{\otimes t}| \le \delta^2$ is plotted at the top and the corresponding number of stabilizer states sampled in the sparsified decomposition with i.i.d. (solid curve) and correlated sampling (dashed curve) is plotted at the bottom. ψ is generated using the SPARSIFY algorithm and the norm is calculated using the FASTNORM algorithm of [4] (using 1000 random stabilizer states to calculate the relative error). [Insets: The difference between the i.i.d. sampling and the correlated sampling curves.]

The statistical distribution of sparsified decompositions from independent sampling and correlated sampling are compared over 1000 numerical runs in Figure 3. The expected value of the state generated by correlated sampling is closer to the desired state (middle of Figure 3). As a result, the standard deviation of the norm of the states generated by correlated sampling is larger (bottom of Figure 3) denoting poorer convergence, as expected. Hence, it is advantageous to use correlated sampling when $\delta^2 \gg (\xi_{\pi/4}^t - t)^{-1}$ to obtain the same convergence probability as independent sampling does at $\delta^2 \gg \xi_{\pi/4}^{-t}$.

At small t a small-number effect occurs since the number of stabilizer states used in correlated sampling is set to the nearest multiple of t greater than or equal to $(\xi_\phi^t - \gamma)\delta^{-2}$ in practice. As a result, at small t more states are sampled than required and this produces a lower expectation value and standard deviation than expected. The factor of $-1/2^{\sim 0.02t}$ in Eq. 17 also reduces the standard deviation at low t.

This method can be extended to produce higher powers of t in γ and thereby improve performance further. In the

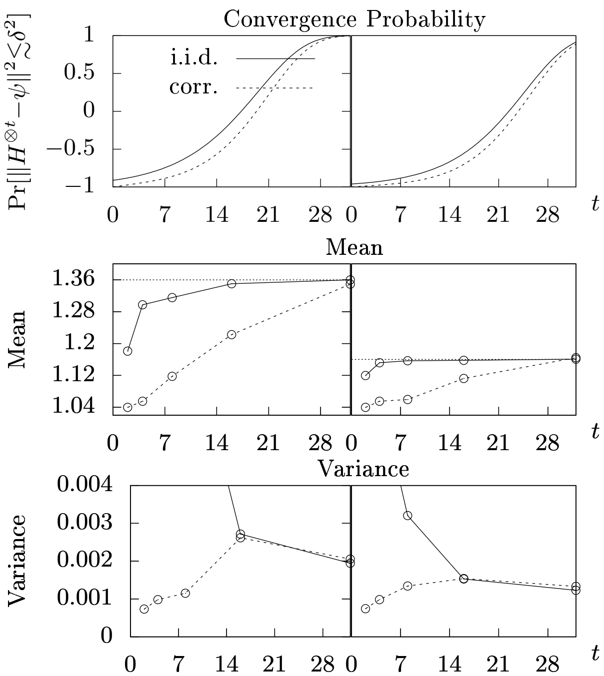

FIG. 3. Plots are for additive error (left) $\delta = 0.6$ and (right) $\delta = 0.4$ over 1000 runs. At the top is plotted the convergence probability lower bound (or sparsification tail bound) given by Eq. 17 for i.i.d. sampling (solid curve) and correlated sampling (dashed curve). In the middle is plotted the mean of $|\langle\psi|\psi\rangle|^2$ for both i.i.d. and correlated sampling to δ error ($1+\delta^2$ is denoted by the dotted horizontal line). At the bottom is plotted the standard deviation of $|\langle\psi|\psi\rangle|^2$ of the sparsified samples, which can be interpreted as a measure of the convergence probability. The standard deviation converges more slowly for large t under correlated sampling than under independent sampling. This agrees with the requirement that $\delta^2 \gg \xi^{-t}$ and $\delta^2 \gg (\xi^t - t)^{-1}$ for i.i.d. and correlated sampling to exhibit $\mathcal{O}(\xi^t\delta^{-2})$ and $\mathcal{O}((\xi^t - t)\delta^{-2})$ scaling, respectively, with the same probability. A small-t effect can be seen where the mean and standard deviation of the correlated samples is lower than that of the i.i.d. samples at $t \lesssim 16$ as explained in the main text. The number of i.i.d. samples and correlated samples generated at particular t-values is shown in Fig. 2.

proof of Theorem 1, the source of the linear power in γ is due to correlated t-wise L_1 sampling; every i.i.d. sampled state is supplemented with t samples with a known relative absolute inner product given by Lemma 1. However, it is easy to show that the number of mutually $\geq t/2$-bitflipped states is larger than t and the number of states given by Lemma 1 is a loose lower bound. The number of supplemented states can be increased to t^m, for $m > 1$, limited by the minimal number $k = (\xi^t_\phi - \gamma)\delta^{-2}$ of stabilizer states needed and the existence of states with the minimum number of mutual bitflips desired. This will add a corresponding power of t^m instead of t in γ. It is also possible to extend γ to higher powers t^m for fixed k by supplementing every i.i.d. sampled state with t samples that have different relative absolute inner products. In both of these cases, doing so would increase the sparsification tail bound of $\Pr\left[\|H^{\otimes t} - \psi\|^2 \leq \langle\Omega|\Omega\rangle - 1 + \delta^2\right]$ further. This would decrease the rate of convergence, and so would require $\delta^2 \gg (\xi^t_{\pi/4} - t^m)^{-1}$ for the improvement in scaling to outperform independent sampling to the same convergence probability. However, it is not clear how many such appropriately bit-flipped supplemental states exist given a bitstring and, therefore, it is not clear how large m of a reduction t^m in γ it is possible to accomplish. We leave this unresolved for future study.

A similar approach will also extend this method of correlated L_1 norm sampling to any of the other diagonal states expressed by Eq. 1. Such a treatment would differ only in that the distribution of t bit-flipped bit strings would be sampled from the non-uniform distribution given by Eq. 1 for $\phi \neq \frac{\pi}{4}$.

Though the stabilizer extent ξ_ϕ of one-, two-, and three-qubit states is multiplicative, general states do not have multiplicative stabilizer extent [8]. This introduces the peculiar notion that L_1 sampling, which is upper bounded by the stabilizer extent (see Lemma 2 in [4]), cannot do better than $\mathcal{O}(2^{\sim 0.228t})$ for the T gate magic state, but that you can always find a more optimal stabilizer decomposition for higher values of t for other states such that their worst-case scaling improves.

The results shown here may resolve this peculiarity. Namely, they show that L_1 sampling is only asymptotically bounded by the stabilizer extent and that, for finite t values, an improvement can be found. This means that the scaling of the L_1 sampling cost of one-, two-, and three- qubit magic states may behave similarly to the scaling of general states.

In conclusion, we show how to lower the finite t scaling cost of the popular SPARSIFY algorithm used in weak simulation of the T+Clifford gateset from $\mathcal{O}(2^{\sim 0.228t}\delta^{-2})$ to $\mathcal{O}((2^{\sim 0.228t} - t)\delta^{-2})$. We accomplish this by replacing its i.i.d. L_1 sampling with correlated L_1 sampling and we numerically demonstrate that this scaling reduction holds after including hidden prefactors polynomial in t. We explain how further reductions in powers of t can be obtained with this method. To our knowledge, this is the first weak simulation algorithm that has lowered this bound's dependence on finite t in the worst-case.

—

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under the Accelerated Research in Quantum Computing program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S.

Department of Energy or the United States Government.

ACKNOWLEDGMENTS

The author thanks Mohan Sarovar for helpful discussions in the process of this research.

AIP PUBLISHING DATA SHARING POLICY

The code that support the findings of this study is openly available in 'https://s3miclassical.com/gitweb/', in the repository 'weak_simulation_stab_extent.git'.

[1] Aram W Harrow and Ashley Montanaro. Quantum computational supremacy. *Nature*, 549(7671):203–209, 2017.

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

[3] Sergey Bravyi, Graeme Smith, and John A Smolin. Trading classical and quantum computational resources. *Physical Review X*, 6(2):021043, 2016.

[4] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Physical review letters*, 116(25):250501, 2016.

[5] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019.

[6] James R Seddon, Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl T Campbell. Quantifying quantum speedups: improved classical simulation from tighter magic monotones. *arXiv preprint arXiv:2002.06181*, 2020.

[7] Hakop Pashayan, Oliver Reardon-Smith, Kamil Korzekwa, and Stephen D Bartlett. Fast estimation of outcome probabilities for quantum circuits. *arXiv preprint arXiv:2101.12223*, 2021.

[8] Arne Heimendahl, Felipe Montealegre-Mora, Frank Vallentin, and David Gross. Stabilizer extent is not multiplicative. *Quantum*, 5:400, 2021.