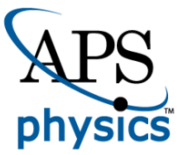

Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates

Sergey Bravyi and David Gosset

Phys. Rev. Lett. **116**, 250501 — Published 20 June 2016

DOI: 10.1103/PhysRevLett.116.250501

Improved classical simulation of quantum circuits dominated by Clifford gates

Sergey Bravyi[1] and David Gosset[2]

[1] *IBM T.J. Watson Research Center, Yorktown Heights NY 10598*
[2] *Walter Burke Institute for Theoretical Physics and Institute for Quantum Information and Matter, California Institute of Technology*

(Dated: May 25, 2016)

We present a new algorithm for classical simulation of quantum circuits over the Clifford+T gate set. The runtime of the algorithm is polynomial in the number of qubits and the number of Clifford gates in the circuit but exponential in the number of T gates. The exponential scaling is sufficiently mild that the algorithm can be used in practice to simulate medium-sized quantum circuits dominated by Clifford gates. The first demonstrations of fault-tolerant quantum circuits based on 2D topological codes are likely to be dominated by Clifford gates due to a high implementation cost associated with logical T-gates. Thus our algorithm may serve as a verification tool for near-term quantum computers which cannot in practice be simulated by other means. To demonstrate the power of the new method, we performed a classical simulation of a hidden shift quantum algorithm with 40 qubits, a few hundred Clifford gates, and nearly 50 T-gates.

The path towards building a large-scale quantum computer will inevitably require verification and validation of small quantum devices. One way to check that such a device is working properly is to simulate it on a classical computer. This becomes impractical at some point because the cost of classical simulation typically grows exponentially with the size of a quantum system. With this fundamental limitation in mind it is natural to ask how well we can do in practice.

Simulation methods which store a complete description of an n-qubit quantum state as a complex vector of size 2^n are limited to a small number of qubits $n \approx 30-40$. For example, a state-of-the art implementation has been used to simulate Shor's factoring algorithm with 31 qubits and roughly half a million gates [1]. Using distributed computation it is possible to simulate 40 qubit circuits [2]. For certain restricted classes of quantum circuits it is possible to do much better [3–7]. Most significantly, the Gottesman-Knill theorem allows efficient classical simulation of quantum circuits composed of gates in the so-called Clifford group [3]. In practice this allows one to simulate such circuits with thousands of qubits [1, 4]. It also means that a quantum computer will need to use gates outside of the Clifford group in order to achieve useful speedups over classical computation. The full power of quantum computation can be recovered by adding a single non-Clifford gate to the Clifford group. A simple choice is the single-qubit $T = |0\rangle\langle 0| + e^{i\pi/4}|1\rangle\langle 1|$ gate. The Clifford+T gate set obtained in this way is a natural instruction set for small-scale fault-tolerant quantum computers based on the surface code [8, 9], and has been at the centre of a recent renaissance in classical techniques for compiling quantum circuits [10–12].

When it comes to realizing a logical (encoded) circuit, non-Clifford gates pose a serious challenge for any fault-tolerant scheme based on 2D stabilizer codes [8, 13] due to the lack of topological protection [14, 15]. Such non-Clifford gates can be implemented fault-tolerantly using special single-qubit resource states known as magic states [16]. The magic states must themselves be prepared using a fault tolerant protocol for "magic state distillation" [16], which is relatively resource intensive. For example, in the case of the surface code, the overhead associated with logical T-gates exceeds that of any logical Clifford gate by orders of magnitude [17, 18]. Thus it is likely that the first logical circuits demonstrated in the lab will be Clifford+T circuits dominated by Clifford gates. In this Letter we propose a new algorithm for classical simulation of such circuits. Our algorithm could therefore serve as a verification tool for near-term quantum computers.

Let us now state our results. A Clifford+T quantum circuit of length m acting on n qubits is a unitary operator $U = U_m \cdots U_2 U_1$, where each U_j is a one- or two-qubit gate from the set $\{H, S, T, CNOT\}$, where

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$$

and $CNOT = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X$ is the controlled-NOT gate. We shall write $m = c + t$, where c is the number of Clifford gates $(H, S, CNOT)$ and t is the number of T-gates also known as the T-count. Applying U to the initial state $|0^{\otimes n}\rangle$ and measuring some fixed output register $Q_{out} \subseteq \{1, \ldots, n\}$ in the $0,1$-basis generates a random bit string x of length $w = |Q_{out}|$. A string x appears with probability

$$P_{out}(x) = \langle 0^{\otimes n}|U^\dagger \Pi(x) U|0^{\otimes n}\rangle, \tag{1}$$

where $\Pi(x)$ projects Q_{out} onto the basis state $|x\rangle$ and acts trivially on the remaining qubits.

Our main result is a classical algorithm for sampling the output string x from a distribution which is ϵ-close to P_{out} with respect to the L_1-norm. The algorithm has runtime

$$\tau = O\left(w(w+t)(c+t) + w(n+t)^3 + 2^{\gamma t} t^3 w^4 \epsilon^{-5}\right), \tag{2}$$

where

$$\gamma \leq -2\log_2(\cos(\pi/8)) \approx 0.228 \tag{3}$$

is a constant that depends on the implementation details. Note that the runtime scales polynomially in all parameters except for the T-count. We expect the algorithm to be practical when the size of the output register w is small and the precision ϵ is not too small. For example, assuming that the circuit outputs a single bit ($w = 1$), ϵ is a fixed constant, and $t \leq n \leq c$, the runtime becomes

$$\tau = O(n^3 + ct + 2^{\gamma t} t^3).$$

The algorithm can be divided into independent subroutines with a runtime $O(t^3)$ each and thus supports a large amount of parallelism. We provide pseudocode for the main steps of the algorithm in the Supplemental Material [19].

Since the simulation runtime is likely to be dominated by the term exponential in t, one may wish to minimize the exponent γ in Eq. (2). This exponent is related to the stabilizer rank [20] of t-qubit tensor product states $|A^{\otimes t}\rangle$, where $|A\rangle$ is a "magic state"

$$|A\rangle = 2^{-1/2}(|0\rangle + e^{i\pi/4}|1\rangle).$$

Recall that a t-qubit state is called a stabilizer state if it has the form $V|0^{\otimes t}\rangle$, where V is a quantum circuit composed of Clifford gates. Stabilizer states form an overcomplete basis in the Hilbert space of t qubits. Let $\chi_t(\delta)$ be the smallest integer χ such that $|A^{\otimes t}\rangle$ can be approximated with an error at most δ by a linear combination of χ stabilizer states (here the approximating state $|\psi\rangle$ should satisfy $|\langle A^{\otimes t}|\psi\rangle|^2 \geq 1-\delta$). The runtime scaling in Eq. (2) holds for any exponent γ such that $\chi_t(\delta) = O(2^{\gamma t})$ for any constant $\delta > 0$ and all sufficiently large t. For simplicity here we assumed that the precision parameter ϵ in Eq. (2) is a constant. Below we propose a systematic method of finding approximate stabilizer decompositions of $|A^{\otimes t}\rangle$ which yields an upper bound $\chi_t(\delta) = O(2^{\gamma t}\delta^{-1})$, where $\gamma \approx 0.228$, see Eq. (3). We conjecture that this upper bound is tight.

We implemented our classical sampling algorithm in MATLAB and used it to simulate a class of benchmark quantum circuits on $n = 40$ qubits, with a few hundred Clifford gates, and T-count $t \leq 48$. Specifically, we simulated a quantum algorithm which solves the hidden shift problem [21] for non-linear Boolean functions [22]. An instance of the hidden shift problem is defined by a pair of oracle functions $f, f' : \mathbb{F}_2^n \to \{\pm 1\}$ and a hidden shift string $s \in \mathbb{F}_2^n$. It is promised that f is a bent (maximally non-linear) function, that is, the Hadamard transform of f takes values ± 1. It is also promised that f' is the shifted version of the Hadamard transform of f, that is,

$$f'(x \oplus s) = 2^{-n/2} \sum_{y \in \mathbb{F}_2^n} (-1)^{x \cdot y} f(y) \quad \text{for all } x \in \mathbb{F}_2^n. \tag{4}$$

Here $\oplus$ stands for the bit-wise XOR. The goal is to learn the hidden shift s by making as few queries to f and f' as possible. The classical query complexity of this problem is known to be linear in n, see Theorem 8 of Ref. [22]. In the quantum setting, f and f' are given as diagonal n-qubit unitary operators O_f and $O_{f'}$ such that $O_f|x\rangle = f(x)|x\rangle$ and $O_{f'}|x\rangle = f'(x)|x\rangle$ for all $x \in \mathbb{F}_2^n$. A quantum algorithm can learn s by making a single query to each of these oracles, as can be seen from the identity [22]

$$|s\rangle = U|0^{\otimes n}\rangle, \quad U \equiv H^{\otimes n} O_{f'} H^{\otimes n} O_f H^{\otimes n}. \tag{5}$$

This hidden shift problem is ideally suited for our benchmarking task for two reasons. First, the algorithm produces a deterministic output, i.e., the output is a computational basis state $|s\rangle$ for some n-bit string s. Because of this we achieve the most favorable runtime scaling in Eq. (2) since each bit of s can be learned by calling the sampling algorithm with a single-qubit output register ($w = 1$) and a constant statistical error ϵ. Second, the T-count of the algorithm can be easily controlled by choosing a suitable bent function. Indeed, the non-oracle part of the algorithm consists only of Hadamard gates. We show that for a large class of bent functions f (from the so-called Maiorana-McFarland family) the oracles O_f and $O_{f'}$ can be constructed using Clifford gates and only a few T-gates.

The numerical simulations were performed for two randomly generated instances of the hidden shift problem with $n = 40$ qubits. For each of these instances we simulated the quantum circuit for the hidden shift algorithm, i.e., the circuit implementing the unitary U described above. The T-counts of the two simulated circuits are $t = 40$ and $t = 48$ respectively. Since the hidden shift s is known beforehand, we are able to verify correctness of the simulation. Our results are presented in Fig. 1. As one can see from the plots, the output probability distribution of each qubit has most of its weight at the corresponding value of the hidden shift bit. Only the output probabilities for qubits $21, 22, \ldots, 40$ are shown because our algorithm perfectly recovered the first half of the hidden shift bits $1, 2, \ldots, 20$. This perfect recovery occurs due to the special structure of the chosen bent functions. Further implementation details can be found in Section IV of the Supplementary Material [19].

Let us now describe two main ingredients of our sampling algorithm. The first ingredient is a subroutine for estimating the norm of a linear combination of stabilizer states. It takes as input a t-qubit state $|\phi\rangle$, a target error parameter $\epsilon > 0$ and a failure probability p_f. The state $|\phi\rangle$ is given as a linear combination of χ stabilizer states,

$$|\phi\rangle = \sum_{a=1}^{\chi} z_a |\phi_a\rangle, \quad \phi_a \in \mathcal{S}_t.$$

Here $\mathcal{S}_t$ is the set of all t-qubit stabilizer states. The

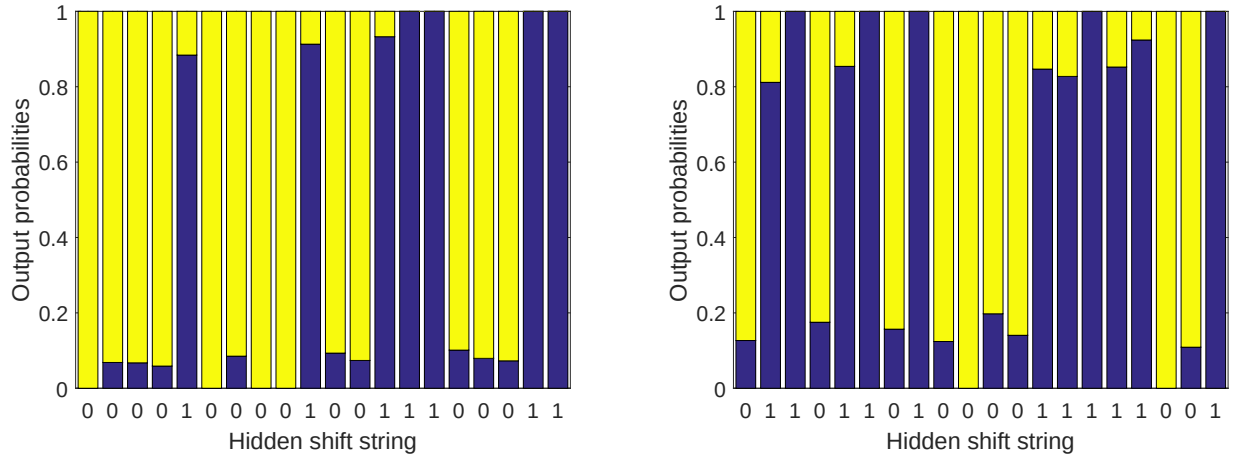

FIG. 1. Output single-qubit probability distributions obtained by a classical simulation of the hidden shift quantum algorithm on $n = 40$ qubits. Only one half of all qubits are shown (qubits $21, 22, \ldots, 40$). The final state of the algorithm is $|s\rangle = U|0^{\otimes n}\rangle$, where s is the hidden shift string to be found and U is a Clifford+T circuit with the T-count $t = 40$ (left) and $t = 48$ (right). In both cases the circuit U contains a few hundred Clifford gates. For each qubit the probability of measuring '1' in the final state is indicated in blue. The x-axis labels indicate the correct hidden shift bits. The entire simulation took several hours on a laptop computer.

subroutine computes a real number ξ which, with probability at least $1-p_f$, approximates the norm $\||\phi\rangle\|^2$ with relative error ϵ. It has running time $O(\chi t^3 \epsilon^{-2} p_f^{-1})$. This improves upon the brute force method which has complexity $O(\chi^2 t^3)$. The key idea is to approximate $\||\phi\rangle\|^2$ by computing inner products between $|\phi\rangle$ and randomly chosen stabilizer states. Let $|\theta\rangle \in \mathcal{S}_t$ be a random stabilizer state drawn from the uniform distribution. Define expectation values

$$M_2 \equiv \mathbb{E}_\theta |\langle\theta|\phi\rangle|^2 \quad \text{and} \quad M_4 \equiv \mathbb{E}_\theta |\langle\theta|\phi\rangle|^4.$$

The set $\mathcal{S}_t$ is known to be a 2-design [28]. This implies that one may compute M_2 and M_4 by pretending that $|\theta\rangle$ is drawn from the Haar measure. Standard formulas for the integrals over the unit sphere yield

$$M_2 = \frac{\||\phi\rangle\|^2}{d} \quad \text{and} \quad M_4 = \frac{2\||\phi\rangle\|^4}{d(d+1)}, \quad \text{where} \quad d \equiv 2^t. \tag{6}$$

Suppose $|\theta_1\rangle, \ldots, |\theta_L\rangle \in \mathcal{S}_t$ are random independent stabilizer states. Define a random variable

$$\xi = \frac{d}{L} \sum_{i=1}^{L} |\langle\theta_i|\phi\rangle|^2. \tag{7}$$

From Eq. (6) one infers that the expected value of ξ is $\bar{\xi} = \mathbb{E}(\xi) = \||\phi\rangle\|^2$ and the standard deviation of ξ is

$$\sigma = \sqrt{d^2 L^{-1}(M_4 - M_2^2)} = \sqrt{\frac{d-1}{d+1}} L^{-1/2} \||\phi\rangle\|^2.$$

For large t one has $\sigma \approx L^{-1/2}\||\phi\rangle\|^2$. By the Chebyshev inequality, $\Pr\Big[\, |\xi - \bar{\xi}| \geq p_f^{-1/2}\sigma \,\Big] \leq p_f$. Thus

$$(1-\epsilon)\||\phi\rangle\|^2 \leq \xi \leq (1+\epsilon)\||\phi\rangle\|^2$$

with probability at least $1 - p_f$ provided that $L = p_f^{-1}\epsilon^{-2}$. The inner product between any t-qubit stabilizer states can be computed classically in time $O(t^3)$, see Refs. [20, 24]. The inner product $\langle\theta_i|\phi\rangle = \sum_{a=1}^{\chi} z_a \langle\theta_i|\phi_a\rangle$ in Eq. (7) can be computed in time $O(\chi t^3)$ since $|\theta_i\rangle$ and $|\phi_a\rangle$ are stabilizer states of t qubits. Thus we compute an approximation to $\||\phi\rangle\|^2$ in time $O(\chi t^3 \epsilon^{-2} p_f^{-1})$. We anticipate that the above norm estimation method can be generalized to stabilizer states of qudits of prime dimension [25] and fermionic Gaussian states [26].

The second ingredient of our simulation algorithm is a method for computing approximate stabilizer decompositions of $|A^{\otimes t}\rangle$. The magic state $|A\rangle$ is equivalent to a state $|H\rangle \equiv \cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle$ modulo Clifford gates and a global phase, $|A\rangle = e^{i\pi/8} HS^\dagger |H\rangle$. Thus it suffices to consider approximate stabilizer decompositions of $|H^{\otimes t}\rangle$. We have the identity

$$|H^{\otimes t}\rangle = \frac{1}{(2\nu)^t} \sum_{x \in \mathbb{F}_2^t} |\tilde{x}_1 \otimes \tilde{x}_2 \otimes \ldots \otimes \tilde{x}_t\rangle \tag{8}$$

where $|\tilde{0}\rangle \equiv |0\rangle$, $|\tilde{1}\rangle \equiv H|0\rangle = 2^{-1/2}(|0\rangle + |1\rangle)$, and $\nu \equiv \cos(\pi/8)$. The right-hand side of Eq. (8) is a uniform superposition of 2^t non-orthogonal stabilizer states labeled by elements of the vector space $\mathbb{F}_2^t$. We construct an approximation $|\psi\rangle$ which is a uniform superposition of states $|\tilde{x}_1 \otimes \tilde{x}_2 \otimes \ldots \otimes \tilde{x}_t\rangle$ over a linear subspace of $\mathbb{F}_2^t$. The dimension k of this subspace is chosen to be the unique positive integer satisfying $4 \geq 2^k \nu^{2t} \delta \geq 2$, where δ is the error tolerance. For any k-dimensional subspace $\mathcal{L}$ of $\mathbb{F}_2^t$ we define a normalized state

$$|\mathcal{L}\rangle = \frac{1}{\sqrt{2^k Z(\mathcal{L})}} \sum_{x \in \mathcal{L}} |\tilde{x}_1 \otimes \tilde{x}_2 \otimes \cdots \otimes \tilde{x}_t\rangle \tag{9}$$

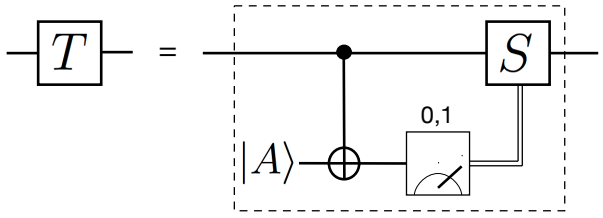

FIG. 2. The T-gate gadget. The Clifford gate S is classically controlled by the measurement outcome.

where $Z(\mathcal{L}) \equiv \sum_{x\in\mathcal{L}} 2^{-|x|/2}$. Here $|\cdot|$ denotes the Hamming weight of a bit string. A simple computation shows that $|\mathcal{L}\rangle$ approximates $|H^{\otimes t}\rangle$ with error

$$\delta(\mathcal{L}) \equiv 1 - |\langle H^{\otimes t}|\mathcal{L}\rangle|^2 = 1 - \frac{2^k \nu^{2t}}{Z(\mathcal{L})}. \tag{10}$$

The error $\delta(\mathcal{L})$ can be computed in time $O(t2^k)$ since $Z(\mathcal{L})$ contains 2^k terms. In Section III of the Supplementary Material we show that by choosing $O(1/\delta)$ k-dimensional subspaces $\mathcal{L}$ uniformly at random we obtain at least one subspace $\mathcal{L}^\star$ such that $\delta(\mathcal{L}^\star) \le \delta$ with high probability. We conclude that $|\langle\psi|A^{\otimes t}\rangle|^2 \ge 1-\delta$, where $|\psi\rangle \equiv (HS^\dagger)^{\otimes t}|\mathcal{L}^\star\rangle$ is a linear combination of $\chi = 2^k = O(\nu^{-2t}\delta^{-1})$ stabilizer states. Computing the approximation $|\psi\rangle$ takes time $O(\nu^{-2t}t\delta^{-2})$. We will see that this is negligible compared with the overall runtime Eq. (2) of the sampling algorithm.

We are now ready to describe the algorithm for sampling from a distribution ϵ-close to P_{out}. For simplicity here we restrict our attention to the case when the output register consists of a single qubit ($w = 1$). We first transform the Clifford+T circuit to be simulated by replacing each T-gate by a certain well-known gadget [27], shown in Fig. 2, that contains only Clifford gates and a $0,1$-measurement. The S gate is classically controlled by the measurement outcome. The gadget consumes one copy of the magic state $|A\rangle$. This gives an equivalent 'gadgetized' circuit consisting of Clifford gates and t single-qubit measurements, acting on a non-stabilizer initial state that contains t copies of $|A\rangle$. Let V_y be the Clifford circuit on $n+t$ qubits corresponding to measurement outcomes described by a t-bit string $y = y_1 y_2 \dots y_t$. Each gadget with $y_j = 0$ contributes a CNOT gate to V_y, whereas each gadget with $y_j = 1$ contributes a CNOT and the S-gate to V_y. Thus V_y contains $c + t + |y|$ gates. Since the gadgetized circuit is equivalent to the original Clifford+T circuit, we have

$$P_{out}(x) = \frac{\langle 0^{\otimes n}\otimes A^{\otimes t}|V_y^\dagger(\Pi(x)\otimes|y\rangle\langle y|)V_y|0^{\otimes n}\otimes A^{\otimes t}\rangle}{\langle 0^{\otimes n}\otimes A^{\otimes t}|V_y^\dagger(I_n\otimes|y\rangle\langle y|)V_y|0^{\otimes n}\otimes A^{\otimes t}\rangle}, \tag{11}$$

for any measurement outcomes y. Let $|\psi\rangle$ be a linear combination of $\chi = O(\nu^{-2t}\delta^{-1})$ stabilizer states constructed above such that $|\langle\psi|A^{\otimes t}\rangle|^2 \ge 1-\delta$. Replacing $|A^{\otimes t}\rangle$ by its approximation $|\psi\rangle$ in Eq. (11) we are led to consider a distribution

$$P_{out}^y(x) = \frac{\langle 0^{\otimes n}\otimes\psi|V_y^\dagger(\Pi(x)\otimes|y\rangle\langle y|)V_y|0^{\otimes n}\otimes\psi\rangle}{\langle 0^{\otimes n}\otimes\psi|V_y^\dagger(I_n\otimes|y\rangle\langle y|)V_y|0^{\otimes n}\otimes\psi\rangle}. \tag{12}$$

This distribution will in general depend on y since $|\psi\rangle$ is not exactly equal to $|A^{\otimes t}\rangle$. In Section II of the Supplementary Material we show that

$$\left\| \frac{1}{2^t}\sum_{y\in\{0,1\}^t} P_{out}^y(x) - P_{out}(x)\right\|_1 = O(\epsilon)$$

provided that $\delta = O(\epsilon^2)$. This shows that we may approximately sample from P_{out} (with error $O(\epsilon)$) by first selecting a t-bit string y uniformly at random and then approximately sampling from P_{out}^y (with error $O(\epsilon)$). It remains to show how to approximately sample from P_{out}^y for a fixed y. Since the gadgetized circuit V_y contains only Clifford gates we may use the standard Gottesman-Knill theorem to compute t-qubit stabilizer groups $\mathcal{G},\mathcal{H}$ and integers u, v such that

$$\langle 0^{\otimes n}\otimes\psi|V_y^\dagger(\Pi(0)\otimes|y\rangle\langle y|)V_y|0^{\otimes n}\otimes\psi\rangle = 2^{-u}\langle\psi|\Pi_{\mathcal{G}}|\psi\rangle \tag{13}$$

$$\langle 0^{\otimes n}\otimes\psi|V_y^\dagger(\Pi(1)\otimes|y\rangle\langle y|)V_y|0^{\otimes n}\otimes\psi\rangle = 2^{-v}\langle\psi|\Pi_{\mathcal{H}}|\psi\rangle \tag{14}$$

where $\Pi_{\mathcal{G}}, \Pi_{\mathcal{H}}$ are projectors onto the codespace of stabilizer codes defined by $\mathcal{G},\mathcal{H}$. This computation, which is described in more detail in Sections I,II of the Supplementary Material, takes time

$$\tau_1 = O(t(c+t) + (n+t)^3).$$

Since we are considering the case where the output string x is a single bit, the output probability distribution is $\{P_{out}^y(0), 1 - P_{out}^y(0)\}$, where

$$P_{out}^y(0) = \frac{2^{-u}\langle\psi|\Pi_{\mathcal{G}}|\psi\rangle}{2^{-v}\langle\psi|\Pi_{\mathcal{H}}|\psi\rangle + 2^{-u}\langle\psi|\Pi_{\mathcal{G}}|\psi\rangle} \tag{15}$$

We compute the expectation values in Eq. (15) with a small relative error using the norm estimation subroutine described above. Indeed, since the projector $\Pi_{\mathcal{G}}$ maps stabilizer states to stabilizer states, one can represent $\Pi_{\mathcal{G}}|\psi\rangle$ as a linear combination of $\chi = O(\nu^{-2t}\epsilon^{-2})$ stabilizer states. Thus one can estimate $\langle\psi|\Pi_{\mathcal{G}}|\psi\rangle = \|\Pi_{\mathcal{G}}|\psi\rangle\|^2$ with a relative error $O(\epsilon)$ and a failure probability $O(\epsilon)$ in time

$$\tau_2 = O(\chi t^3\epsilon^{-3}) = O(\nu^{-2t}t^3\epsilon^{-5}).$$

Let $\xi = 2^{-u}\langle\psi|\Pi_{\mathcal{G}}|\psi\rangle(1\pm\epsilon)$ and $\xi' = 2^{-v}\langle\psi|\Pi_{\mathcal{H}}|\psi\rangle(1\pm\epsilon)$ be the resulting approximations. The final step in the algorithm is to sample a bit from the probability distribution $\{p_0, 1-p_0\}$ where $p_0 = \xi/(\xi+\xi')$ (cf. Eq. (15)).

The approximation guarantees for ξ, ξ' ensure that this distribution is $O(\epsilon)$-close to P^y_{out}. The total runtime of the algorithm is $\tau_1 + \tau_2$ from which we recover the $w = 1$ case of Eq. (2).

Whereas here we focused on the case $w = 1$, in Section II of the Supplementary Material we describe the simulation algorithm for arbitrary w. Although this algorithm can be used for sampling from the output distribution with a small statistical error, in general it cannot accurately compute individual probabilities of the output distribution. In the Supplementary Material we also present a different algorithm which uses similar techniques to compute the output probabilities $P_{out}(x)$ with a relative error ϵ.

ACKNOWLEDGMENTS

DG acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028). SB thanks Alexei Kitaev for helpful discussions and comments.

[1] D. Wecker and K. M. Svore, preprint arXiv:1402.4467 (2014).
[2] M. Smelyanskiy, N. P. D. Sawaya, and A. Aspuru-Guzik, preprint arXiv:1601.07195 (2016).
[3] D. Gottesman, preprint quant-ph/9807006 (1998).
[4] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).
[5] I. Markov and Y. Shi, SIAM J. on Comp. **38**, 963 (2008).
[6] M. Van den Nest, Quant. Inf. Comp. **10**, 0258 (2010).
[7] H. Pashayan, J. Wallman, and S. Bartlett, Phys. Rev. Lett. **115**, 070501 (2015).
[8] S. Bravyi and A. Kitaev, preprint quant-ph/9811052 (1998).
[9] A. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Phys. Rev. A **86**, 032324 (2012).
[10] V. Kliuchnikov, D. Maslov, and M. Mosca, Quant. Inf. and Comp. **13**, 607 (2013).
[11] P. Selinger, Quant. Inf. and Comp. **15**, 159 (2015).
[12] N. J. Ross and P. Selinger, preprint arXiv:1403.2975 (2014).
[13] H. Bombin and M. A. Martin-Delgado, Physical Review Letters **97**, 180501 (2006).
[14] S. Bravyi and R. König, Phys. Rev. Lett. **110**, 170503 (2013).
[15] F. Pastawski and B. Yoshida, Phys. Rev. A **91**, 012305 (2015).
[16] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).
[17] A. Fowler, S. Devitt, and C. Jones, Scientific reports **3**, 1939 (2013).
[18] C. Jones, preprint arXiv:1310.7290 (2013).
[19] See Supplemental Material at [URL] for further details on the implementation of our algorithm and numerical simulations..
[20] S. Bravyi, G. Smith, and J. Smolin, preprint arXiv:1506.01396 (2015).
[21] W. van Dam, S. Hallgren, and L. Ip, in *Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms* (2003), pp. 489–498.
[22] M. Rötteler, in *Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms* (2010), pp. 448–457.
[28] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80**, 012304 (2009).
[24] H. J. García, I. Markov, and A. Cross, Quant. Inf. and Comp. **14**, 683 (2014).
[25] E. Hostens, J. Dehaene, and B. De Moor, Phys. Rev. A **71**, 042315 (2005).
[26] S. Bravyi, Quant. Inf. and Comp. **5**, 216 (2005).
[27] X. Zhou, D. W. Leung, and I. L. Chuang, Phys. Rev. A **62**, 052316 (2000).
[28] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80**, 012304 (2009).
[29] P. Hayden, D. Leung, P. W. Shor, and A. Winter, Comm. in Math. Phys. **250**, 371 (2004).
[30] C. Jones, Phys. Rev. A **87**, 022328 (2013).
[31] M. Amy, D. Maslov, M. Mosca, and M. Roetteler, Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on **32**, 818 (2013), 1206.0758.
[32] D. Gosset, V. Kliuchnikov, M. Mosca, and V. Russo, Quant. Inf. and Comp. **14**, 1261 (2014).
[33] K.-U. Schmidt, Information Theory, IEEE Transactions on **55**, 5803 (2009).
[34] M. Araújo, *Classification of quadratic forms*, http://www.math.ist.utl.pt/~ggranja/manuel.pdf (2011).
[35] J. Dehaene and B. De Moor, Phys. Rev. A **68**, 042318 (2003).
[36] H. García-Ramírez, Ph.D. thesis, The University of Michigan (2014).