Simulation of quantum circuits by low-rank stabilizer decompositions

Sergey Bravyi[1], Dan Browne[2], Padraic Calpin[2], Earl Campbell[3], David Gosset[1,4], and Mark Howard[3]

[1]IBM T.J. Watson Research Center, Yorktown Heights NY 10598

[2]Department of Physics and Astronomy, University College London, London, UK

[3]Department of Physics and Astronomy, University of Sheffield, Sheffield, UK

[4]Department of Combinatorics & Optimization and Institute for Quantum Computing, University of Waterloo, Waterloo, Canada

Recent work has explored using the stabilizer formalism to classically simulate quantum circuits containing a few non-Clifford gates. The computational cost of such methods is directly related to the notion of *stabilizer rank*, which for a pure state ψ is defined to be the smallest integer χ such that ψ is a superposition of χ stabilizer states. Here we develop a comprehensive mathematical theory of the stabilizer rank and the related approximate stabilizer rank. We also present a suite of classical simulation algorithms with broader applicability and significantly improved performance over the previous state-of-the-art. A new feature is the capability to simulate circuits composed of Clifford gates and arbitrary diagonal gates, extending the reach of a previous algorithm specialized to the Clifford+T gate set. We implemented the new simulation methods and used them to simulate quantum algorithms with 40-50 qubits and over 60 non-Clifford gates, without resorting to high-performance computers. We report a simulation of the Quantum Approximate Optimization Algorithm in which we process superpositions of $\chi \sim 10^6$ stabilizer states and sample from the full n-bit output distribution, improving on previous simulations which used $\sim 10^3$ stabilizer states and sampled only from single-qubit marginals. We also simulated instances of the Hidden Shift algorithm with circuits including up to 64 T gates or 16 CCZ gates; these simulations showcase the performance gains available by optimizing the decomposition of a circuit's non-Clifford components.

Contents

1 Introduction

It is widely believed that universal quantum computers cannot be efficiently simulated by classical probabilistic algorithms. This belief is partly supported by the fact that state-of-the-art classical simulators employing modern supercomputers are still limited to a few dozens of qubits [17, 30, 48, 52]. At the same time, certain quantum information processing tasks do not require computational universality. For example, quantum error correction based on stabilizer codes and Pauli noise models [27] only requires quantum circuits composed of Clifford gates and Pauli measurements–which can be easily simulated classically for thousands of qubits using the Gottesman-Knill theorem [2, 6]. Furthermore, it is known that Clifford circuits can be promoted to universal quantum computation when provided with a plentiful supply of some computational primitive outside the stabilizer operations, such as a non-Clifford gate or magic state [12]. This raises the possibility of simulating quantum circuits with a large number of qubits and few non-Clifford gates. Aaronson and Gottesman [2] were the first to propose a classical simulation method covering this situation, with a runtime that scales polynomially with the number of qubits and Clifford gate count but exponentially with the number of non-Clifford gates. This early work is an intriguing proof of principle but with a very large exponent, limiting potential applications.

Recent algorithmic improvements have helped tame this exponential scaling by significantly decreasing the size of the exponent. A first step was made by Garcia, Markov and Cross [25, 26], who proposed and studied the decomposition of states into a superposition of stabilizer states. Bravyi, Smith and Smolin [14] formalized this into the notion of stabilizer rank. The stabilizer rank $\chi(\psi)$ of a pure state ψ is defined as the smallest integer χ such that ψ can be expressed as a superposition of χ stabilizer states. It can be thought of as a measure of computational non-classicality analogous the Schmidt rank measure of entanglement. In particular, $\chi(\psi)$ quantifies the simulation cost of stabilizer operations (Clifford gates and Pauli measurements) applied to the initial state ψ.

It is known that stabilizer operations augmented with preparation of certain single-qubit "magic states" become computationally universal [12]. In particular, any quantum circuit composed of Clifford gates and m gates $T = |0\rangle\langle 0| + e^{i\pi/4}|1\rangle\langle 1|$ can be implemented by stabilizer operations acting on the initial state $|\psi\rangle = |T\rangle^{\otimes m}$, where $|T\rangle \propto |0\rangle + e^{i\pi/4}|1\rangle$. Thus the stabilizer rank $\chi(T^{\otimes m})$ provides an upper bound on the simulation cost of Clifford+T circuits with m T-gates. The authors of Ref. [14] used a numerical search method to compute the stabilizer rank $\chi(T^{\otimes m})$ for $m \leq 6$ finding that $\chi(T^{\otimes 6}) = 7$. The numerical search becomes impractical for $m > 6$ and one instead works with suboptimal decompositions by breaking m magic states up into blocks of six or fewer qubits. This yields a classical simulator of Clifford+T circuits running in time $2^{0.48m}$ with certain polynomial prefactors [11]. More recently, Ref. [11] introduced an approximate version of the stabilizer rank and a method of constructing approximate stabilizer decomposition of the magic states $|T\rangle^{\otimes m}$. This led to a simulation algorithm with runtime scaling as $2^{0.23m}$ that samples the output distribution of the target circuit with a small statistical error. In practice, it can simulate

single-qubit measurements on the output state of Clifford+T circuits with $m \leq 50$ on a standard laptop [11]. A similar class of simulation methods uses Monte Carlo sampling over quasiprobability distributions, where the distribution can be over either a discrete phase space [20, 47, 56], over the class of stabilizer states [32] or over stabilizer operations [7]. These quasiprobability methods are a natural method for simulating noisy circuits but for pure circuits they appear to be slower than simulation methods based on stabilizer rank.

Here we present a more general set of tools for finding exact and approximate stabilizer decompositions as well as improved simulation algorithms based on such decompositions. A central theme throughout this paper is generalizing the results of Refs. [11, 14] beyond the Clifford+T setting. While Clifford+T is a universal gate set, it requires several hundred T gates to synthesize an arbitrary single qubit gate to a high precision (e.g. below 10^{-10} error). Therefore, it would be impractical to simulate such gates using the Clifford+T framework. We achieve significant improvements in the simulation runtime by branching out to more general gate sets including arbitrary-angle Z-rotations and CCZ gates. Furthermore, we propose more efficient subroutines for simulating the action of Clifford gates and Pauli measurements on superpositions of $\chi \gg 1$ stabilizer states. In practice, this enables us to perform simulations in the regime $\chi \sim 10^6$ with about 50 qubits on a laptop computer improving upon $\chi \sim 10^3$ simulations reported in Ref. [11]. The table provided below summarizes new simulation methods, simulation tasks addressed by each method, and the runtime scaling.

Method		Gate set	Simulation	Stabilizer decomposition	Runtime
Gadget based	Fixed sample	Clifford+T	Strong	Exact	$2^{0.48m}$
	Random sample	Clifford+T	Weak	Approx	$2^{0.23m}$
Sum over Cliffords	Norm estimation	Clifford+$R(\theta)$	Weak	Approx	$\left(\cos(\theta/2) + \tan\left(\frac{\pi}{8}\right)\sin(\theta/2)\right)^{2m}$
	Metropolis	Clifford+$R(\theta)$	Weak	Approx	Varies

Figure 1: Summary of new simulation methods. For simplicity, here we restrict the attention to quantum circuits composed of Clifford gates and diagonal single-qubit gates $R(\theta) = \mathrm{diag}(1, e^{i\theta})$. The T-gate can be obtained as a special case $T = R(\pi/4)$. We consider strong and weak simulation tasks where the goal is to estimate a single output probability (with a small multiplicative error) and sample the output probability distribution (with a small statistical error) respectively. The runtime scales exponentially with the non-Clifford gate count m and polynomially with the number of qubits and the Clifford gate count. For simplicity, here we ignore the polynomial prefactors. For a detailed description of our simulation methods, see Section 2.3.

On the theory side, we establish some general properties of the approximate stabilizer rank. Our main tool is a Sparsification Lemma that shows how to convert a dense stabilizer decomposition of a given target state (that may contain all possible stabilizer states) to a sparse decomposition that contains fewer stabilizer states. The lemma generalizes the method of random linear codes introduced in Ref. [11] in the context of Clifford+T circuits. It allows us to obtain sparse stabilizer decompositions for the output state of more general quantum circuits directly without using magic state gadgets. Combining the Sparsification Lemma and convex duality arguments, we relate the approximate stabilizer rank of a state ψ to a stabilizer fidelity $F(\psi)$ defined as the maximum overlap between ψ and stabilizer states. Central to these calculations is a new quantity called Stabilizer Extent, which quantifies, in an operationally relevant way, how non-stabilizer a state is. We give necessary and sufficient conditions under which the stabilizer fidelity is multiplicative under the tensor product. Finally, we propose a new strategy for proving lower bounds on the

stabilizer rank of the magic states which uses the machinery of ultra-metric matrices [41, 45].

As a main application of our simulation algorithms we envision verification of noisy intermediate-size quantum circuits [49] in the regime when a brute-force classical simulation may be impractical [3, 35, 44]. For example, a quantum circuit composed of Clifford gates and single-qubit Z-rotations with angles $\theta_1, \ldots, \theta_m$ can be efficiently simulated using our methods in the regime when only a few of the angles θ_a are non-zero or if all the angles θ_a are small in magnitude, see Section 2.3.2. By fixing the Clifford part of the circuit and varying the rotation angles θ_a one can therefore interpolate between the regimes where the circuit output can and cannot be verified classically. From the experimental perspective, single-qubit Z-rotations are often the most reliable elementary operations [43]. Thus one should expect that the circuit output fidelity should not depend significantly on the choice of the angles θ_a.

The next section provides a more detailed overview of our results.

2 Main results

Recall that the Clifford group is a group of unitary n-qubit operators generated by single-qubit and two-qubit gates from the set $\{H, S, CX\}$. Here H is the Hadamard gate, $S = |0\rangle\langle 0| + i|1\rangle\langle 1|$ is the phase shift gate, and CX=CNOT is the controlled-X gate. Stabilizer states are n-qubit states of the form $|\phi\rangle = U|0^n\rangle$, where U is a Clifford operator. We also use X_j, Y_j, Z_j to denote Pauli operators acting on the j-th qubit. Below we also make use of the stabilizer formalism, and refer the unfamiliar reader to the existing literature [46].

2.1 Tools for constructing low-rank stabilizer decompositions

In this section we summarize our results pertaining to the stabilizer rank and describe methods of decomposing a state into a superposition of stabilizer states. A reader interested only in the application for simulation of quantum circuits may wish to proceed to Sections 2.2, 2.3.

Definition 1 (**Exact stabilizer rank, χ [14]**). *Suppose ψ is a pure n-qubit state. The exact stabilizer rank $\chi(\psi)$ is the smallest integer k such that ψ can be written as*

$$|\psi\rangle = \sum_{\alpha=1}^{k} c_\alpha |\phi_\alpha\rangle, \tag{1}$$

for some n-qubit stabilizer states ϕ_α and some complex coefficients c_α.

By definition, $\chi(\psi) \geq 1$ for all ψ and $\chi(\psi) = 1$ iff ψ is a stabilizer state.

Definition 2 (**Approximate stabilizer rank, χ_δ [11]**). *Suppose ψ is a pure n-qubit state such that $\|\psi\| = 1$. Let $\delta > 0$ be a precision parameter. The approximate stabilizer rank $\chi_\delta(\psi)$ is the smallest integer k such that $\|\psi - \psi'\| \leq \delta$ for some state ψ' with exact stabilizer rank k.*

Note that this definition differs slightly from the one from Ref. [11] which is based on the fidelity. Our first result provides an upper bound on the approximate stabilizer rank.

Theorem 1 (**Upper bound on χ_δ**). *Let ψ be a normalized n-qubit state with a stabilizer decomposition $|\psi\rangle = \sum_{\alpha=1}^{k} c_\alpha |\phi_\alpha\rangle$ where $|\phi_\alpha\rangle$ are normalized stabilizer states and $c_\alpha \in \mathbb{C}$. Then*

$$\chi_\delta(\psi) \leq 1 + \|c\|_1^2/\delta^2. \tag{2}$$

Here $\|c\|_1 \equiv \sum_{\alpha=1}^{k} |c_\alpha|$.

We note that the stabilizer decomposition $|\psi\rangle = \sum_{\alpha=1}^{k} c_\alpha |\phi_\alpha\rangle$ in the statement of the theorem does not have to be optimal. For example, it may include all stabilizer states. The proof of the theorem is provided in Section 5.1. It is constructive in the sense that it provides a method of calculating a state ψ' which is a superposition of $\chi' \approx \delta^{-2}\|c\|_1^2$ stabilizer states such that $\|\psi' - \psi\| \leq \delta$. Such a state ψ' is obtained using a randomized sparsification method. It works by sampling χ' stabilizer states ϕ_α from the given stabilizer decomposition of ψ at random with

probabilities proportional to $|c_\alpha|$. The state ψ' is then defined as a superposition of the sampled states ϕ_α with equal weights, see the Sparsification Lemma and related discussion in Section 5.2. The theorem motivates the following definition.

Definition 3 (**Stabilizer Extent, ξ**). *Suppose ψ is a normalized n-qubit state. Define the stabilizer extent $\xi(\psi)$ as the minimum of $\|c\|_1^2$ over all stabilizer decompositions $|\psi\rangle = \sum_{\alpha=1}^{k} c_\alpha |\phi_\alpha\rangle$ where ϕ_α are normalized stabilizer states.*

The theorem immediately implies that

$$\chi_\delta(\psi) \leq 1 + \xi(\psi)/\delta^2. \tag{3}$$

While it is difficult to compute or prove tight bounds for the exact or approximate stabilizer rank, we find that $\xi(\psi)$ is a more amenable quantity that can be calculated for many states ψ relevant in the context of quantum circuit simulation. In particular, we prove

Proposition 1 (**Multiplicativity of Stabilizer Extent**). *Let $\{\psi_1, \psi_2, \ldots, \psi_L\}$ be any set of states such that each state ψ_j describes a system of at most three qubits. Then*

$$\xi(\psi_1 \otimes \psi_2 \otimes \ldots \otimes \psi_L) = \prod_{j=1}^{L} \xi(\psi_j). \tag{4}$$

This shows that the upper bound of Theorem 1 is multiplicative under tensor product in the case of few-qubit states. It remains open whether ξ is multiplicative on arbitrary collections of states.

The proof of Proposition 1 is provided in Section 6.4. It uses the fact that standard convex duality provides a characterization of ξ in terms of the following quantity.

Definition 4 (**Stabilizer Fidelity, F**). *The stabilizer fidelity, $F(\psi)$, of a state ψ is*

$$F(\psi) = \max_\phi |\langle \phi | \psi \rangle|^2, \tag{5}$$

where the maximization is over all normalized stabilizer states ϕ.

Proposition 1 is obtained as a consequence of new results concerning multiplicativity of the stabilizer fidelity. In particular, we apply the classification of entanglement in three-partite stabilizer states [13] to derive conditions for the multiplicativity of $F(\psi)$. More precisely, we define a set of quantum states $\mathcal{S}$ which we call *stabilizer aligned* such that $F(\phi \otimes \psi) = F(\phi)F(\psi)$ whenever $\phi, \psi \in \mathcal{S}$. A state ψ is called stabilizer aligned if the overlap between ψ and any stabilizer projector of rank 2^k is at most $2^{k/2}F(\psi)$. Remarkably, the set of stabilizer aligned states is closed under tensor product, that is $\phi \otimes \psi \in \mathcal{S}$ whenever $\phi, \psi \in \mathcal{S}$. Moreover, we show that the stabilizer fidelity is not multiplicative for all states $\phi \notin \mathcal{S}$. That is, for any $\phi \notin \mathcal{S}$ there exists a state ψ such that $F(\phi \otimes \psi) > F(\phi)F(\psi)$. In that sense, our results provide necessary and sufficient conditions under which the stabilizer fidelity is multiplicative.

Proposition 1 enables computation of $\xi(\psi)$ if ψ is a tensor product of few-qubit states (that involve at most three qubits). We now describe another large subclass of states ψ relevant for quantum circuit simulation for which we are able to compute ξ. To describe these states, recall that any diagonal t-qubit gate V can be performed using a state-injection gadget that contains only stabilizer operations and consumes an ancillary state $|V\rangle = V|+\rangle^{\otimes t}$ (see the discussion in Section 2.3 and Figure 2). Here and below $|+\rangle \equiv (|0\rangle + |1\rangle)/\sqrt{2}$. The gadget also involves a computational basis measurement over t qubits. Let $x \in \{0,1\}^t$ be a string of measurement outcomes. The desired gate V is performed whenever $x = 0^t$. However, given some other outcome $x \neq 0^t$, the gadget implements a gate $V_x = C_x V$ where

$$C_x = \prod_{j\,:\,x_j=1} V X_j V^\dagger,$$

is the required correction, where X_j is the Pauli X operator acting on the jth qubit. A special class of unitaries are those where the correction C_x is always a Clifford operator. In this case a unitary gate V is equivalent to the preparation of the ancillary state $|V\rangle$ modulo stabilizer operations. This motivates the following definition.

Definition 5 (Clifford magic states). *Let V be a diagonal t-qubit unitary such that $VX_jV^\dagger$ is a Clifford operator for all j. Such unitary V is said to belong to the 3rd level of the Clifford hierarchy (see e.g. Ref. [28]). The ancillary state $|V\rangle \equiv V|+\rangle^{\otimes t}$ is called a Clifford magic state.*

For example, $|T\rangle^{\otimes m}$ is a Clifford magic state for any integer m. Note that in general the set of Clifford magic states is closed under tensor product.

Proposition 2. *Let ψ be a Clifford magic state. Then $\xi(\psi) = F(\psi)^{-1}$.*

The proof of Proposition 2 is provided in Section 5.3 where it is extended to a slightly broader class of ψ.

We note that $|T^{\otimes m}\rangle$ is a Clifford magic state *and* a product state and so either Proposition 1 or Proposition 2 could be used along with Eq. (3) to upper bound its approximate stabilizer rank. In this way one can easily reproduce the upper bound obtained in Ref. [11], namely,

$$\chi_\delta(T^{\otimes m}) \leq O\left(\delta^{-2}\cos(\pi/8)^{-2m}\right). \tag{6}$$

This stands in sharp contrast with the best known lower bound $\chi(T^{\otimes m}) = \Omega(m^{1/2})$ established in Ref. [14]. It should be expected that the stabilizer rank (either exact or approximate) of the magic states $T^{\otimes m}$ grows exponentially with m in the limit $m \to \infty$. Indeed, the polynomial scaling of $\chi_\delta(T^{\otimes m})$ with m for a suitably small constant δ, or $\chi(T^{\otimes m})$, would entail complexity theoretic heresies such as BQP=BPP, or P=NP [1]. Remarkably, we have no techniques for proving unconditional super-polynomial lower bounds. Here we made partial progress by solving a simplified problem where stabilizer decompositions of $T^{\otimes m}$ are restricted to certain product states. For this simplified setting we prove a tight lower bound on the approximate stabilizer rank of $T^{\otimes m}$ matching the upper bound of Ref. [11]. To state our result it is more convenient to work with the magic state $|H\rangle = \cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle$ which is equivalent to $|T\rangle$ modulo Clifford gates. Ref. [11] showed that $|H^{\otimes m}\rangle$ admits an approximate stabilizer decomposition $|H^{\otimes m}\rangle \approx \sum_{\alpha=1}^{k} c_\alpha|\phi_\alpha\rangle$ where $k \sim \cos(\pi/8)^{-2m}$ and ϕ_α are product stabilizer states of the form

$$|\tilde{x}\rangle = |\tilde{x}_1\rangle \otimes \cdots \otimes |\tilde{x}_m\rangle \quad \text{where} \quad |\tilde{0}\rangle = |0\rangle \quad \text{and} \quad |\tilde{1}\rangle = |+\rangle. \tag{7}$$

Here $x_i \in \{0,1\}$. These are the stabilizer states that achieve the maximum overlap with $|H^{\otimes m}\rangle$, see Ref. [11]. Here we prove the following lower bound.

Proposition 3. *Suppose $S \subseteq \{0,1\}^m$ is an arbitrary subset and ψ is an arbitrary linear combination of states $|\tilde{x}\rangle$ as in (7) with $x \in S$ such that $\|\psi\| = 1$. Then*

$$|S| \geq |\langle H^{\otimes m}|\psi\rangle|^2 \cdot \cos(\pi/8)^{-2m}. \tag{8}$$

The proof of this result which is given in Section 5.4 makes use of the machinery of ultra-metric matrices [41, 45]. We hope that these techniques may lead to further progress on lower bounding the stabilizer rank.

We conclude this section by summarizing our results pertaining to the exact stabilizer rank. Prior work focused exclusively on finding the stabilizer rank of m-fold tensor products of magic state $|T\rangle$. A surprising and counter-intuitive result of Ref. [14] is that for small number of magic states ($m \leq 6$) the stabilizer rank $\chi(T^{\otimes m})$ scales linearly with m. Meanwhile, $\chi(T^{\otimes m})$ is expected to scale exponentially with m in the limit $m \to \infty$. Using a numerical search we observed a sharp jump from $\chi(T^{\otimes 6}) = 7$ to $\chi(T^{\otimes 7}) = 12$ indicating a transition from the linear to the exponential scaling at $m = 7$. This poses the question of whether other magic states have a linearly scaling stabilizer rank (until some critical m is reached) or if $|T\rangle$ is an exceptional state due to its special symmetries. Here we show that the linear scaling for small m is a generic feature.

Theorem 2 (Upper bound on χ). *Let ψ be an n-qubit state and then for all $m \leq 5$ we have*

$$\chi(\psi^{\otimes m}) \leq \binom{2^n + m - 1}{m} \tag{9}$$

where the round brackets denote the binomial coefficient.

[1] By simulating a postselective quantum circuit one could solve 3-SAT using a polynomial number of T-gates, see e.g., Ref. [33].

For example, this result shows that for any diagonal single-qubit unitary V the associated magic state $|V\rangle$ obeys $\chi(|V\rangle^{\otimes m}) \leq m+1$ for $m \leq 5$. For larger m, an exponential scaling is expected. The proof of Theorem 2 (given in Section 5.1) exploits well-known properties of the symmetric subspace and a recently established fact that n-qubit stabilizer states form a 3-design [38, 57].

2.2 Subroutines for manipulating low-rank stabilizer decompositions

Suppose U is a quantum circuit acting on n qubits. We consider a classical simulation task where the goal is to sample a bit string $x \in \{0,1\}^n$ from the probability distribution $P_U(x) = |\langle x|U|0^n\rangle|^2$ with a small statistical error.

Suppose we are given an approximate stabilizer decomposition of a state $U|0^n\rangle$:

$$\||U|0^n\rangle - |\psi\rangle\| \leq \delta, \qquad |\psi\rangle = \sum_{\alpha=1}^{k} b_\alpha U_\alpha |0^n\rangle \tag{10}$$

for some coefficients b_α and some Clifford circuits U_α. In Section 4 we give algorithms for the following tasks. These algorithms are the main subroutines used in our quantum circuit simulators.

(a) Sample $x \in \{0,1\}^n$ from the probability distribution

$$P(x) = \frac{|\langle x|\psi\rangle|^2}{\||\psi\|^2}. \tag{11}$$

(b) Estimate the norm $\||\psi\|^2$ with a small multiplicative error.

Note that if δ is small then $P(x)$ approximates the true output distribution $P_U(x) = |\langle x|U|0^n\rangle|^2$ with a small error. Indeed, Eq. (10) gives $\|P - P_U\|_1 \leq O(\delta)$.

The tasks (a,b) are closely related. Using the chain rule for conditional probabilities one can reduce the sampling task to estimation of marginal probabilities of $P(x)$. Any marginal probability can be expressed as $\|\Pi\psi\|^2/\|\psi\|^2$, where Π is a tensor product of projectors $|0\rangle\langle 0|$, $|1\rangle\langle 1|$, and the identity operators. Note that such projectors map stabilizer states to stabilizer states. Thus $\Pi|\psi\rangle$ admits a stabilizer decomposition with k terms that can be easily obtained from Eq. (10). Accordingly, task (a) reduces to a sequence of norm estimations for low-rank stabilizer superpositions, see Section 4.3 for details.

Section 4.1 describes a fast Clifford simulator that transforms a stabilizer state $U_\alpha|0^n\rangle$ into a certain canonical form which we call a CH-form. It is analogous to the stabilizer tableaux [2] but includes information about the global phase of a state. This allows us to simulate each circuit U_α in the superposition Eq. (10) independently without destroying information about the relative phases. Our C++ implementation of the simulator performs approximately 5×10^6 Clifford gates per second for $n = 64$ qubits on a laptop computer.

Section 4.2 describes a heuristic algorithm for the task (a). We construct a Metropolis-type Markov chain such that $P(x)$ is the unique steady distribution of the chain (under mild additional assumptions). We show how to implement each Metropolis step in time $O(kn)$. Unfortunately, the mixing time of the chain is generally unknown.

Section 4.3 gives an algorithm for the task (b). It exploits the fact that the inner product $\langle\phi|\phi'\rangle$ between two n-qubit stabilizer states ϕ, ϕ' can be computed exactly in time $O(n^3)$, see Ref. [11, 25]. We adapt this inner product algorithm to the CH-form of stabilizer states in Section 4.3. The naive method of computing the norm relies on the identity $\||\psi\|^2 = \sum_{\alpha,\beta=1}^{k} b_\alpha^* b_\beta \langle\phi_\alpha|\phi_\beta\rangle$, where $|\phi_\alpha\rangle = U_\alpha|0^n\rangle$. Evaluating all cross terms using the inner product algorithm would take time $O(k^2n^3)$ which is impractical for large k. Instead, Ref. [11] proposed a method of estimating, rather than evaluating, the norm. It works by computing inner products between ψ and random stabilizer states drawn from the uniform distribution. This method has runtime $O(kn^3)$ offering a significant speedup in the relevant regime of large rank decompositions. Here we propose an improved version of this norm estimation method combining both conceptual and implementation improvements. The new version of the norm estimation subroutine achieves approximately 50X speedup compared with Ref. [11].

Section 4.3 also describes a rigorous algorithm for the task (a) based on the norm estimation and the chain rule for conditional probabilities. It has runtime $O(kn^6)$ which quickly becomes impractical. However, if our goal is to sample only w bits from $P(x)$, the runtime is only $O(kn^3w^3)$. Thus the sampling method based on the norm estimation may be practical for small values of w.

2.3 Simulation algorithms

Here we describe how to combine ingredients from previous sections to obtain classical simulation algorithms for quantum circuits. We consider a circuit

$$U = D_m V_m D_{m-1} V_{m-1} \dots D_1 V_1 D_0 \tag{12}$$

acting on input state $|0^n\rangle$, where $\{D_j\}$ are Clifford circuits and $\{V_j\}$ are non-Clifford gates. We discuss three different methods: gadget-based simulation (using either a fixed-sample or random-sample method as described below) and sum-over-Cliffords simulation.

Let us first summarize the simulation cost of different methods. The gadget-based methods from Refs. [11, 14] can be used to simulate quantum circuits Eq. (12) where $\{V_j\}$ are single-qubit T gates. Using the (random-sample) gadget-based method, the asymptotic cost of sampling from a distribution δ-close in total variation distance to the output distribution $P_U(x) = |\langle x|U|0^n\rangle|^2$ is

$$\tilde{O}\left(\chi_\delta\left(|T^{\otimes m}\rangle\right)\right) \leq \tilde{O}\left(\delta^{-2}\xi\left(|T^{\otimes m}\rangle\right)\right) = \tilde{O}\left(\delta^{-2}\left(\cos(\pi/8)\right)^{-2m}\right), \tag{13}$$

where we used Theorem 1 and Proposition 1, and the $\tilde{O}$-notation suppresses a factor polynomial in m, n, and $\log(\delta^{-1})$, see Ref. [11] for details.

We will see how the gadget-based approach can be applied in a slightly more general setting where the circuit contains diagonal gates from the third level of the Clifford hierarchy. Then we introduce the sum-over-Cliffords simulation method which can be applied much more generally. The cost of δ-approximately sampling from the output distribution P_U for the circuit Eq. (12) using the sum-over-Cliffords method can be upper bounded as

$$\tilde{O}\left(\delta^{-2}\prod_{j=1}^{m}\xi(V_j)\right) \tag{14}$$

where the definition of ξ is extended to unitary matrices in a natural way (see below for a formal definition). For example, if each non-Clifford gate is a single-qubit diagonal rotation of the form $V_j = R(\theta_j) = e^{-i(\theta_j/2)Z}$ with $\theta_j \in [0, \pi/2)$ then we will see that $\xi(V_j) = \xi(V_j|+\rangle)$ and the simulation cost is

$$\tilde{O}\left(\delta^{-2}\prod_{j=1}^{m}\xi(V_j|+\rangle)\right) = \tilde{O}\left(\delta^{-2}\prod_{j=1}^{m}\left(\cos(\theta_j/2) + \tan(\pi/8)\sin(\theta_j/2)\right)^2\right).$$

In the case $\theta_j = \pi/4$ where all non-Cliffords are T gates, we see that the sum-over-Cliffords method achieves the same asymptotic cost Eq. (13) as the gadget-based method from Ref. [11]. However the sum-over-Cliffords method is generally preferred because it is simpler to implement and may be slightly faster, as it manipulates stabilizer states of fewer qubits.

2.3.1 Gadget-based methods

We begin by reviewing the gadget-based methods for simulating circuits expressed over the Clifford+T gate set. A gadget-based simulation directly emulates the operation of a quantum computer that can implement Clifford operations and has access to a supply of magic states.

It is well known that one can perform such a gate on a quantum computer using a state-injection gadget with classical feedforward dependent on measurement outcomes. In particular, a t-qubit gate V can be implemented by a gadget consuming a magic state $|V\rangle = V|+^t\rangle$, see Fig. 2 for an example. Let $x \in \{0,1\}^t$ be the measurement outcome. The gadget implements the desired gate V whenever $x = 0^t$. Otherwise, if $x \neq 0^t$, the gadget implements a gate $V_x = C_x^\dagger V$ where C_x is the required correction. If V is in the third level of the Clifford hierarchy, the correction C_x is always

a Clifford operator and $|V\rangle$ is a Clifford magic state (recall Definition 5). Formally, postselecting on outcome $x = 0^t$ gives

$$V|\psi\rangle = 2^{t/2}(\mathbb{1} \otimes \langle 0|^{\otimes t})C'|\psi\rangle|V\rangle, \tag{15}$$

where $C' = \left(\prod_{a=1}^{t} \mathrm{CNOT}_{a,a+t}\right)$ is a Clifford unitary.

Now let U from Eq. (12) be the full circuit to be simulated and suppose V_j is a diagonal t_j-qubit gate. Write $\tau = t_1 + t_2 \ldots + t_m$. If we replace each non-Clifford gate with the corresponding state-injection gadget we obtain a "gadgetized" circuit with $n + \tau$ qubits acting on input state $|0^n\rangle|V_1\rangle|V_2\rangle \ldots |V_m\rangle$. The gadgetized circuit contains τ extra single-qubit measurements and Clifford gates. If we postselect the measurement outcomes on 0^τ we obtain an identity (cf. Eq. (15))

$$U|0^n\rangle = 2^{\tau/2}(\mathbb{1} \otimes \langle 0|^{\otimes \tau})C|0^n\rangle|\Psi\rangle \qquad |\Psi\rangle = |V_1\rangle|V_2\rangle \ldots |V_m\rangle \tag{16}$$

where C is an $n+\tau$-qubit Clifford unitary and we have collected together all of the required magic states into the τ-qubit state Ψ. We see a renormalisation factor $2^{\tau/2}$ is required to account for post-selection.

Eq. (16) shows that the output state $U|0^n\rangle$ of interest has exact stabilizer rank equal to that of the magic state Ψ, i.e., $\chi(U|0^n\rangle) = \chi(\Psi)$. Indeed, starting from an exact stabilizer decomposition of $|\Psi\rangle$, we can apply $(\mathbb{1} \otimes \langle 0|^{\otimes \tau})C$ to each stabilizer state in the decomposition and renormalize to obtain an exact stabilizer decomposition of the output state $U|0^n\rangle$. Once we have computed an exact stabilizer decomposition of $U|0^n\rangle$ we may use the subroutines from Section 2.2 to simulate the quantum computation. For example we may sample from the output distribution P_U or compute a given output probability $P_U(x)$. This was the approach taken in Ref. [14] and here we call this a fixed-sample gadget-based simulator since it postselects on a fixed single measurement outcome.

Note that in the fixed-sample method one must use an exact (rather than approximate) stabilizer decomposition of the resource state Ψ. Indeed, in a fixed-sample simulation if $|\Psi_\delta\rangle$ approximates $|\Psi\rangle$ up to an error δ then the simulation error could be amplified to $2^{\tau/2}\delta$ when substituting in Eq. (16).

The random-sample gadget-based simulation method is a different approach that allows us to use approximate stabilizer decompositions within this framework. Here one selects the post-selected measurement outcome $x \in \{0,1\}^\tau$ uniformly at random. However, now we have some measurement outcomes other than $x = 0^\tau$ and so have to account for corrections C_x. Clifford corrections are straightforwardly simulated and this is ensured provided each non-Clifford gate V_j in the circuit is diagonal in the computational basis and contained in the third level of the Clifford hierarchy (e.g., the T gate and CCZ gate). This guarantees that the simulation consuming an approximate magic state $|\Psi_\delta\rangle$ achieves an average-case simulation error $O(\delta)$, see Ref. [11] for details.

An important distinction between the two gadget-based methods is that the random-sample method allows one to sample from a probability distribution which approximates P_U but–unlike the fixed-sample method– in general cannot be used to obtain an accurate estimate of an individual output probability $P_U(x)$.

2.3.2 Sum-over-Cliffords method

Let U be the quantum circuit Eq. (12) to be simulated. We shall construct a sum-over-Cliffords decomposition

$$U = \sum_j c_j K_j \tag{17}$$

where each K_j is a unitary Clifford operator and c_j are some coefficients. This gives

$$U|0^n\rangle = \sum_j c_j K_j |0^n\rangle. \tag{18}$$

Applying Theorem 1 one can approximate $U|0^n\rangle$ within any desired error δ by a superposition of stabilizer states ψ that contains

$$k \approx \delta^{-2}\|c\|_1^2 \tag{19}$$

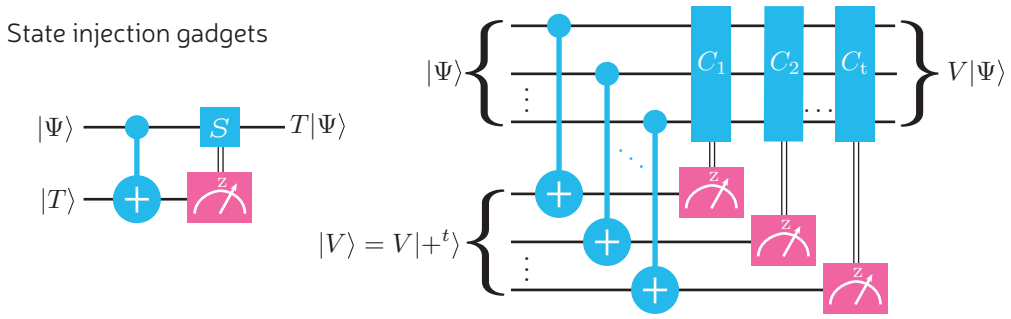


Figure 2: State injection gadgets for single-qubit T gate and general multi-qubit phase gate V. A correction unitary $VX_jV^\dagger$ is required whenever measurement j registers a "1" outcome. If all corrections are Clifford then gadgets can be deployed with no additional resource requirements.

terms. In this way we can compute an approximate stabilizer decomposition ψ satisfying

$$\||U|0^n\rangle - |\psi\rangle\| \leq \delta, \qquad |\psi\rangle = \sum_{\alpha=1}^{k} b_\alpha U_\alpha |0^n\rangle, \tag{20}$$

for some coefficients b_α and some Clifford circuits U_α. Using the methods summarized in the previous section we can then sample from the distribution $P(x) = |\langle x|\psi\rangle|^2$ which δ-approximates the output distribution P_U. In particular, one can use either the heuristic Metropolis sampling technique or the rigorous algorithm using norm estimation, which has runtime upper bounded as $O(kn^6)$.

The sum-over-Cliffords decomposition Eq. (17) of U can be obtained by combining decompositions of the constituent non-Clifford gates. If $V_p = \sum_j c_j^{(p)} K_j^{(p)}$ for $p = 1, 2, \ldots, m$, then substituting in Eq. (12) gives

$$U = \sum_{j_1,\ldots,j_m} \left(\prod_{p=1}^{m} c_{j_p}^{(p)} \right) D_m K_{j_m}^{(m)} D_{m-1} \ldots D_1 K_{j_1}^{(1)} D_0$$

which is of the form Eq. (17) with $\|c\|_1^2 = \prod_{p=1}^m \|c^{(p)}\|_1^2$. This motivates the following generalization of ξ to unitary operators.

Definition 6 (Stabilizer Extent for unitaries, cf. Eq. 14). *Suppose W is a unitary operator. Define $\xi(W)$ as the minimum of $\|c\|_1^2$ over all decompositions $W = \sum_j c_j K_j$ where K_j are Clifford unitaries.*

This implies

$$\xi(U|0^n\rangle) \leq \xi(U) \leq \prod_j \xi(V_j). \tag{21}$$

Thus, given ξ-optimal decompositions of each non-Clifford gate in the circuit, the asymptotic cost of δ-approximately sampling from $P_U(x)$ using the norm estimation algorithm and the sum-over-Cliffords method is $\tilde{O}(k)$, and substituting Eq. (21) in Eq. (19) we recover Eq. (14).

Note that for any gate V_j which acts on $O(1)$ qubits we may compute a ξ-optimal sum-over-Cliffords decomposition in constant time by an exhaustive search. Below we describe decompositions for commonly used non-Clifford gates. We use the following lemma which "lifts" a stabilizer decomposition of the resource state $|V\rangle = V|+^t\rangle$ to a sum-over-Cliffords decomposition of V.

Lemma 1 (Lifting lemma). *Suppose V is a diagonal t-qubit unitary and*

$$V|+^t\rangle = |V\rangle = \sum_j c_j |\phi_j\rangle. \tag{22}$$

Suppose further that $|\phi_j\rangle$ are equatorial stabilizer states so that $|\phi_j\rangle = K_j|+^t\rangle$ where K_j is a diagonal Clifford for all j. Then

$$V = \sum_j c_j K_j, \tag{23}$$

and therefore $\xi(V) \leq ||c||_1^2$. Furthermore, if the equatorial stabilizer decomposition Eq. (22) achieves the optimal value $\|c\|_1^2 = \xi(|V\rangle)$ then $\xi(|V\rangle) = \xi(V)$.

Proof. Since U and $\{K_j\}$ are diagonal in the computational basis we may write

$$V = \sum_x e^{i\theta(x)}|x\rangle\langle x| \qquad K_j = \sum_x e^{i\theta_j(x)}|x\rangle\langle x| \tag{24}$$

where θ, θ_j are functions $\mathbb{F}_2^t \to \mathbb{R}$. For all $x \in \{0,1\}^t$ we have

$$\frac{1}{2^{t/2}} e^{i\theta(x)} = \langle x|V|+^t\rangle = \langle x| \sum_j c_j K_j |+^t\rangle = \frac{1}{2^{t/2}} \sum_j c_j e^{i\theta_j(x)} \tag{25}$$

Combining Eqs. (24,25) and cancelling the factors of $2^{-t/2}$ gives Eq. (23) and the remaining statements of the lemma are immediate corollaries. □

For single-qubit diagonal rotations $R(\theta) = e^{-i(\theta/2)Z}$, we have

$$R(\theta)|+\rangle = (\cos(\theta/2) - \sin(\theta/2))\,|+\rangle + \sqrt{2}\sin(\theta/2)e^{-i\pi/4}S|+\rangle, \tag{26}$$

which is an optimal decomposition with respect to ξ and is similar to Eq. (163). Therefore, we can use the lifting lemma to obtain an optimal decomposition

$$R(\theta) = (\cos(\theta/2) - \sin(\theta/2))\,\mathbb{1} + \sqrt{2}e^{-i\pi/4}\sin(\theta/2)S \tag{27}$$

and conclude

$$\xi(R(\theta)) = \xi(R(\theta)|+\rangle) = (\cos(\theta/2) + \tan(\pi/8)\sin(\theta/2))^2\,. \tag{28}$$

The doubly controlled Z gate (CCZ) is another useful example. In Section 5.3 we show that

$$\begin{aligned} |CCZ\rangle &= \frac{2}{9}(\mathbb{1} + CZ_{1,2}X_3)(\mathbb{1} + CZ_{1,3}X_2)(\mathbb{1} + CZ_{2,3}X_1)|+^3\rangle, \\ &= \frac{2}{9}\big(\mathbb{1} + CZ_{1,2} + CZ_{1,3} + CZ_{2,3} + CZ_{1,2}CZ_{1,3}Z_1 + CZ_{1,2}CZ_{2,3}Z_2 \\ &\quad + CZ_{1,3}CZ_{2,3}Z_3 - CZ_{1,2}CZ_{1,3}CZ_{2,3}Z_1Z_2Z_3\big)|+^3\rangle, \end{aligned} \tag{29}$$

is an optimal decomposition with respect to ξ. Deploying the lifting lemma we have

$$\begin{aligned} CCZ = \frac{2}{9}\big(&\mathbb{1} + CZ_{1,2} + CZ_{1,3} + CZ_{2,3} + CZ_{1,2}CZ_{1,3}Z_1 + CZ_{1,2}CZ_{2,3}Z_2 \\ &+ CZ_{1,3}CZ_{2,3}Z_3 - CZ_{1,2}CZ_{1,3}CZ_{2,3}Z_1Z_2Z_3\big), \end{aligned} \tag{30}$$

and conclude

$$\xi(CCZ) = \xi(|CCZ\rangle) = 16/9. \tag{31}$$

Recall that since this is a Clifford magic state we have $\xi(|CCZ\rangle) = 1/F(|CCZ\rangle)$ and notice that the stabilizer fidelity is achieved by the equatorial stabilizer state $|+^3\rangle$. We remark that the above recipe for an optimal sum-over-Cliffords decomposition can be generalised to any Clifford magic state for which the stabilizer fidelity is achieved by some equatorial stabilizer state.

These optimal sum-over-Cliffords decompositions will be used in the numerics of the following Section.

2.4 Implementation and simulation results

In this section we report numerical results obtained by simulating two quantum algorithms. First, we use the sum-over-Cliffords method to simulate the Quantum Approximate Optimization (QAOA) algorithm due to Farhi et al [22]. This algorithm allows us to explore the performance of our simulator for circuits containing Cliffords and diagonal rotations. This simulation involves $n = 50$ qubits, about 60 non-Clifford gates, and a few hundred Clifford gates. We note that QAOA circuits have been previously used to benchmark classical simulators in Ref. [24]. Secondly, we simulate the Hidden Shift algorithm for bent functions due to Roetteler [51]. This algorithm was also used to benchmark the Clifford+T simulator of Ref. [11] which, in the terminology of the previous section, is a gadget-based simulator where sparsification is achieved via suitable choice of a random linear code. We extend this methodology to a Clifford+CCZ simulator of the same circuits. We also simulate the Hidden Shift circuits using the new Sum-over-Cliffords method wherein sparsification is achieved by appealing to the ξ quantity.

2.4.1 Quantum approximate optimization algorithm

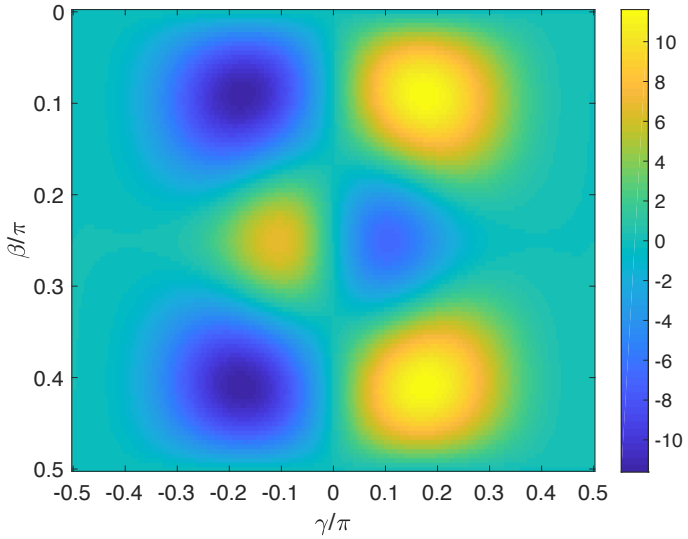


Figure 3: The expected value of the cost function $E(\beta, \gamma)$ computed using the Monte Carlo method by Van den Nest [55]. We consider a randomly generated instance of the Max E3LIN2 problem with $n = 50$ qubits and degree $D = 4$.

Here we consider the Quantum Approximate Optimization Algorithm applied to the Max E3LIN2 problem [22]. The problem is to maximize an objective function

$$C = \frac{1}{2} \sum_{1 \le u < v < w \le n} d_{uvw} z_u z_v z_w$$

that depends on n binary variables $z_1, \ldots, z_n \in \{-1, 1\}$. Here $d_{uvw} \in \{0, \pm 1\}$ are some coefficients. Let

$$m = \sum_{u<v<w} |d_{uvw}|$$

be the number of non-zero terms in C. Let us say that an instance of the E3LIN2 problem has degree D if each variable z_u appears in exactly D terms $\pm z_u z_v z_w$ (depending on the values of n and D there could be one variable that appears in less than D terms).

Following Ref. [22] we consider a family of variational states

$$|\psi_{\beta,\gamma}\rangle = U|0^n\rangle \qquad U = e^{-i\beta B} e^{-i\gamma \hat{C}} H^{\otimes n}$$

where $\beta, \gamma \in \mathbb{R}$ are variational parameters, $B = X_1 + \ldots + X_n$ is the transverse field operator, and $\hat{C}$ is a diagonal operator obtained from C by replacing the variables z_u with the Pauli operators Z_u. The QAOA algorithm attempts to choose β and γ maximizing the expected value of the objective function,

$$E(\beta, \gamma) = \langle \psi_{\beta,\gamma} | \hat{C} | \psi_{\beta,\gamma} \rangle.$$

Once a good choice of β, γ is made, the QAOA algorithm samples $z \in \{-1, 1\}^n$ from a probability distribution $P(z) = |\langle z | \psi_{\beta,\gamma} \rangle|^2$ by preparing the state $|\psi_{\beta,\gamma}\rangle$ on a quantum computer and measuring each qubit of $|\psi_{\beta,\gamma}\rangle$. (In this section we assume that output bits take values ± 1 rather than $0, 1$.) By definition, the expected value of $C(z)$ coincides with $E(\beta, \gamma)$. By generating sufficiently many samples one can produce a string z such that $C(z) \geq E(\beta, \gamma)$, see Ref. [22] for details.

Our numerical results described below were obtained for a single randomly generated instance of the problem with $n = 50$ qubits and degree $D = 4$. We empirically observed that the expected value $E(\beta, \gamma)$ does not depend significantly on the choice of the problem instance for fixed n and D. Since the cost function has a symmetry $C(-z) = -C(z)$, finding the maximum and the minimum values of C are equivalent problems.

A special feature of the QAOA circuits making them suitable for benchmarking classical simulators is the ability to verify that the simulator is working properly. This is achieved by computing the expected value $E(\beta, \gamma)$ using two independent methods and cross checking the final answers. Our first method of computing $E(\beta, \gamma)$ is a classical Monte Carlo algorithm due to Van den Nest [55]. It allows one to compute expected values $\langle \omega | F | \omega \rangle$, where F is an arbitrary sparse Hamiltonian and $|\omega\rangle$ is a so-called computationally tractable state. Let us choose $F = e^{i\beta B} \hat{C} e^{-i\beta B}$ and $|\omega\rangle = e^{-i\gamma \hat{C}} |+^{\otimes n}\rangle$ so that $\langle \omega | F | \omega \rangle = E(\beta, \gamma)$. The algorithm of Ref. [55] allows one to estimate $\langle \omega | F | \omega \rangle$ with an additive error ϵ in time $O(m^4 \epsilon^{-2})$. The plot of $E(\beta, \gamma)$ is shown on Fig. 3.

Our second method of computing $E(\beta, \gamma)$ is the sum-over-Cliffords/Metropolis simulator described in Section 2.3.2. We used this method to simulate the QAOA circuit U defined above. For our choice $n = 50$ and $D = 4$ the unitary $e^{-i\gamma \hat{C}}$ can be implemented by a circuit that contains $m = 66$ Z-rotations $e^{i(\gamma/2)Z}$ and a few hundred Clifford gates. To keep the number of non-Clifford gates sufficiently small we restricted the simulations to the line $\beta = \pi/4$. As can be seen from Fig. 3, this line contains a local maximum and a local minimum of $E(\beta, \gamma)$ (we note that $\beta = \pi/4$ is also the choice made by Farhi et al. [22]). With this choice the cost function is a function of a single parameter γ and we may write

$$E(\gamma) = \langle 0^n | U^\dagger \hat{C} U | 0^n \rangle = \sum_{z \in \{0,1\}^n} P_U(z) C(z).$$

between the "exact" value $E(\gamma)$ computed by the Monte Carlo method and its estimate $E_{sim}(\gamma)$ obtained using the sum-over-Cliffords/Metropolis simulator (while the Monte Carlo method is not perfect, we expect the errors to be negligible for our purposes). While the plot only shows $\gamma \geq 0$, note that due to the symmetry of the cost function $C(z) = -C(-z)$ we have $E(\gamma) = -E(-\gamma)$. The estimate $E_{sim}(\gamma)$ is defined as

$$E_{sim}(\gamma) = \frac{1}{s} \sum_{j=1}^{s} C(z^j), \qquad s = 4 \cdot 10^4$$

where $z^1, \ldots, z^s$ are samples from the distribution $P(z)$ describing the output of the simulator, see Eq. (11). Generating all of the data used to produce Fig. 4a took less than 3 days on a laptop computer, with the most costly data points taking several hours. The number of stabilizer states k used to approximate $U|0^n\rangle$ is shown in Fig. 4b; it was chosen as in Eq. (19) with $\delta \leq 0.15$ for all values of γ. This toy example demonstrates that our algorithm is capable of processing superpositions of $k \sim 10^6$ stabilizer states for $n = 50$ qubits.

2.4.2 The hidden shift algorithm

In this section, we describe the results of simulations applied to a family of quantum circuits that solve the Hidden Shift Problem [53] for non-linear Boolean functions [51]. These circuits are identical to those simulated in [11] and further details of this quantum algorithm and its circuit

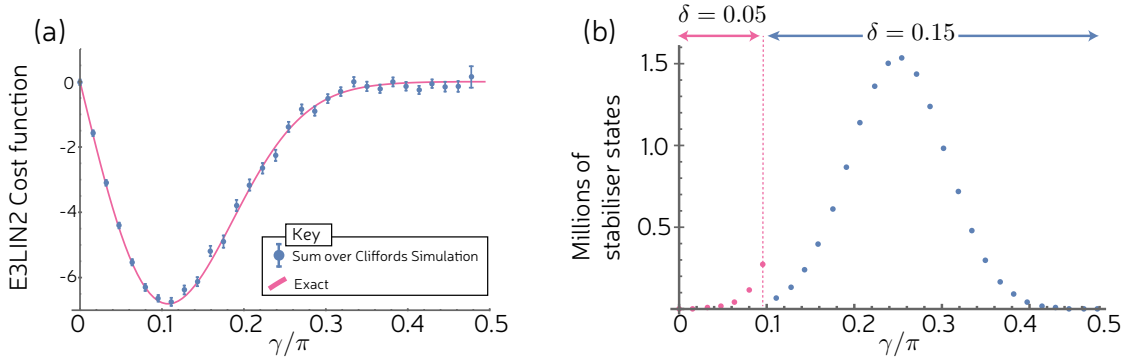


Figure 4: *Classical simulation of the QAOA algorithm:* (*a*) Comparison between $E(\gamma)$ and its estimate $E_{sim}(\gamma)$ obtained using the sum-over-Cliffords/Metropolis simulator. We consider a randomly generated instance of the problem with $n = 50$ qubits and degree $D = 4$. For each data point 10^4 Metropolis steps were performed to approach the steady distribution $P(z)$. The estimate $E_{sim}(\gamma)$ was obtained by averaging the cost function $C(z)$ over a subsequent $s = 4 \cdot 10^4$ samples x from the output distribution of the simulator. Error bars represent the statistical error estimated using the MATLAB code due to Wolff [58] (for estimating errors in Markov chain Monte Carlo data) (*b*) The number of stabilizer states k used by the sum-over-Cliffords simulator was chosen as in Eq. (19) with $\delta = 0.05$ for pink data points and $\delta = 0.15$ for blue data points.

instantiation can be found in Section F of the Supplemental Material of [11]. Briefly, the goal is to learn a hidden shift string $s \in \mathbb{F}_2^n$ by measuring the output state $|s\rangle$ of the circuit U applied to computational basis input $|0^{\otimes n}\rangle$. The number of non-Clifford gates in U can easily be controlled (we may choose any even number of Toffoli gates) and so the exponentially growing overhead in simulation time can be observed.

We will use both the gadget-based method of Section 2.3.1 and the Sum-over-Cliffords method of 2.3.2. Due to the high number of non-Clifford gates the exact stabilizer rank, χ, is prohibitively high and so some sort of sparsification/approximation must be used, leading to χ_δ instead. In principle we could apply the sparsification Lemma 6 in the gadget-based setting, but we prefer to use the random code method of [11] to enable a comparison with that work. The simulation timings in Fig. 5 consist of four trend lines which can be broken down as

- T_{GB}: The gadget-based random code method of [11], wherein each Toffoli gate in U is decomposed in terms of a stabilizer circuit using 4 T gadgets. When a gadgetized version of U uses a total of t $|T\rangle$-type magic states, then $|T^{\otimes t}\rangle$ is approximated by a state $|\mathcal{L}\rangle$ where $\mathcal{L} \subseteq \mathbb{F}_2^t$ is a linear subspace i.e., random code (Compare with Eq. (105)). We then have that $\chi_\delta(|T^{\otimes t}\rangle)$ is the number of vectors in $\mathcal{L}$.

- CCZ_{GB}: The gadget-based random code method of [11], wherein each Toffoli gate in U is implemented via a CCZ gadget (as discussed e.g., in [32]). When gadgetized U uses a total of u $|CCZ\rangle$-type magic states, then $|CCZ^{\otimes u}\rangle$ is approximated by a state $|\mathcal{L}\rangle$ (see Eq. (105)) where $\mathcal{L} \subseteq \mathbb{F}_2^{3u}$ is a linear subspace/random code and $\chi_\delta(|CCZ^{\otimes u}\rangle) = |\mathcal{L}|$.

- T_{SoC}: The Sum-over-Cliffords method outlined in Sec. 4, wherein each Toffoli gate in U is decomposed in terms of a stabilizer circuit using 4 T gates. Each T gate is subsequently decomposed into Clifford gates, $T = c_0 I + c_1 S$, with weightings as in Eq. (27).

- CCZ_{SoC}: The Sum-over-Cliffords method outlined in Sec. 4, wherein each Toffoli gate in U written as CCZ which is subsequently decomposed (optimally in terms of ξ) into Cliffords as in Eq. (30).

The quantity that eventually determines the simulation overhead for both the T-based and CCZ-based schemes is F, the overlap with the closest stabilizer state. Recall $\xi(T) = \xi(|T\rangle) =$

$1/F(|T\rangle)$ and likewise for CCZ. We have

$$F(T) = |\langle +|T\rangle|^2 = \cos(\pi/8)^2 = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.853, \tag{32}$$

$$F(CCZ) = |\langle +^{\otimes 3}|CCZ\rangle|^2 = \left(\frac{3}{4}\right)^2 = \frac{9}{16}. \tag{33}$$

Note that we are using the variable u to denote the number of Toffoli (equivalently CCZ) gates in our Hidden Shift circuit. Using the Random Code method, for a target infidelity Δ we chose a corresponding stabilizer rank 2^k where [11] stipulates

$$\log_2 k_T = \lfloor \log_2 \left(4\cos(\pi/8)^{-8u}/\Delta\right) \rfloor, \tag{34}$$

$$\log_2 k_{CCZ} = \lfloor \log_2 \left(4\left(\tfrac{3}{4}\right)^{-2u}/\Delta\right) \rfloor. \tag{35}$$

Using the Sum-over-Cliffords method, for a target error δ we chose k as in Lemma 6 so that

$$k_T = \left\lfloor \left(\cos(\pi/8)^{-4u}/\delta\right)^2 \right\rfloor, \tag{36}$$

$$k_{CCZ} = \lfloor \left((3/4)^{-u}/\delta\right)^2 \rfloor. \tag{37}$$

In either case, we see that there are significant savings to be had by using CCZ gates/states directly versus breaking them down into 4 T gates/states each. For a fixed precision the scaling with u (number of CCZ gates) goes as

$$T: \quad \left(\frac{1}{\cos \pi/8}\right)^{8u} \approx 2^{0.914u}, \tag{38}$$

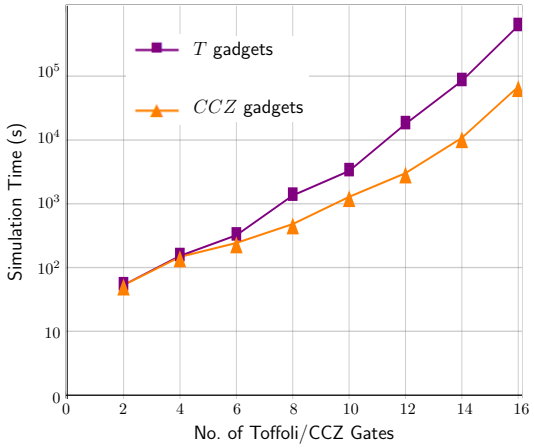
$$\text{vs.} \quad CCZ: \quad \left(\frac{16}{9}\right)^{u} \approx 2^{0.83u}. \tag{39}$$

This is apparent from the different slopes of the T- and CCZ- based versions of the simulations in Fig. 5.

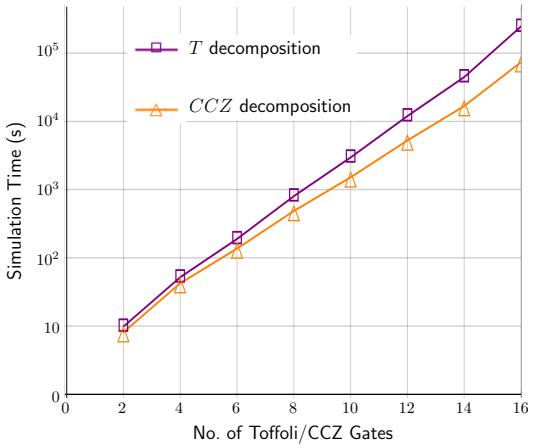
Absolute comparisons between the gadget-based and Sum-over-Cliffords method are complicated by various implementation details and the amount of optimization applied to each (i.e., more in the latter case). Broadly speaking, however, we observe that the Sum-over-Cliffords method is as fast, if not faster, than the gadget-based method. This is true *despite the fact that Sum-over-Cliffords is completely general in its applicability* whereas the gadget-based technique is only applicable for non-Clifford gates from the third level of the Clifford hierarchy (i.e. those with state-injection gadgets having Clifford corrections). Not only can Sum-over-Cliffords handle gates outside the third level, its performance often *improves* in such situations. For example, a circuit with many small-angle rotation gates requires a number, k, of samples that is smaller as the rotation angle moves away from $\pi/4$ i.e., the T case (recall Eq. (27)).

3 Discussion

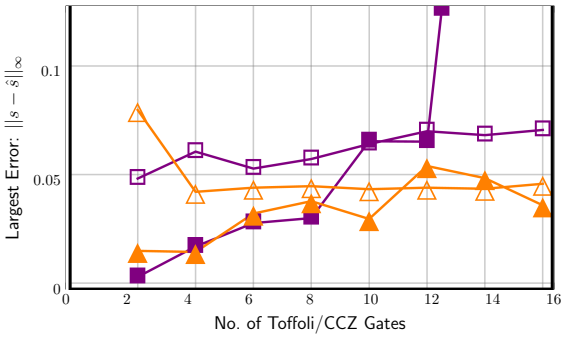
To put our results in a broader context, let us briefly discuss alternative methods for classical simulation of quantum circuits. Vector-based simulators [19, 30, 52] represent n-qubit quantum states by complex vectors of size 2^n stored in a classical memory. The state vector is updated upon application of each gate by performing sparse matrix-vector multiplication. The memory footprint limits the method to small number of qubits. For example, Häner and Steiger [30] reported a simulation of quantum circuits with $n = 45$ qubits and a few hundred gates using a supercomputer with 0.5 petabytes of memory. In certain special cases the memory footprint can be reduced by recasting the simulation problem as a tensor network contraction [1, 8, 40]. Several tensor-based simulators have been developed [17, 39, 48] for geometrically local shallow quantum circuits that include only nearest-neighbor gates on a 2D grid of qubits [9]. These methods enabled simulations of systems with more than 100 qubits [17]. However, it is expected [33] that for general



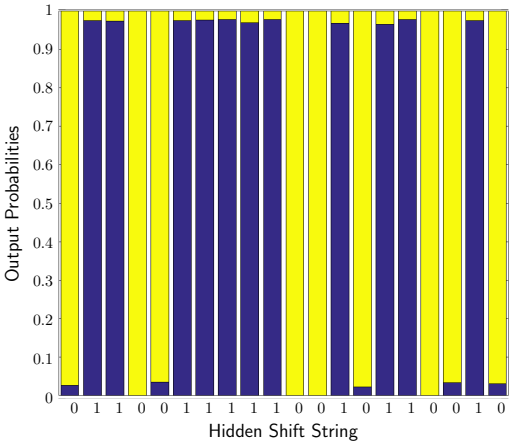
(a) Simulation time for Hidden Shift circuits using the gadget-based random code method from Ref. [11].



(b) Simulation time for Hidden Shift circuits using the Sum-over-Cliffords method from 4.1.



(c) Approximation error between the true hidden shift bitstring, s, and the simulated vector of marginal probabilities, $\hat{s}$, for the simulations in Fig. 5a and 5b. The infinity norm gives the largest discrepancy between any individual bit s_i and the corresponding estimate $\hat{s}_i$. Two outlier data points (filled rectangles) whose coordinates are at $(14, 0.304)$ and $(16, 0.512)$, are omitted from this plot for clarity



(d) Simulated output, $\hat{s}$, versus the true shift string, s, for the case T_{SoC} with 16 Toffoli gates (i.e corresponding to the open rectangle on the right of 5c).

Figure 5: Timings and errors for simulations of 40-qubit Hidden Shift circuits with varying numbers of non-Clifford gates. Every Toffoli gate is either recast as a CCZ gate (via Hadamards on the target) or as a circuit comprising 4 T gates and additional Stabilizer operations ([11]). We fixed precision parameters $\delta = 0.3$ and $\Delta = 0.3$ for the sum-over-Clifford simulations and gadget-based simulations respectively. Simulations were run on Dual Intel Xeon 1.90GHz processors using Matlab.

(geometrically non-local) circuits of size $poly(n)$ the runtime of tensor-based simulators scales as $2^{n-o(n)}$.

In contrast, Clifford simulators described in the present paper are applicable to large-scale circuits without any locality properties as long as the circuit is dominated by Clifford gates. This regime may be important for verification of first fault-tolerant quantum circuits where logical non-Clifford gates are expected to be scarce due to their high implementation cost [23, 34]. Another advantage of Clifford simulators is their ability to sample the output distribution of the circuit (as opposed to computing individual output amplitudes). This is more close to what one would expect from the actual quantum computer. For example, a single run of the heuristic sum-over-Cliffords simulator described in Section 4.2 produces thousands of samples from the (approximate) output distribution. In contrast, a single run of a tensor-based simulator typically computes a single amplitude of the output state. Thus we believe that our techniques extend the reach of classical

simulation algorithms complementing the existing vector- or tensor-based simulators.

A version of the sum-over-Cliffords simulator using the Metropolis sampling method is also publicly available as part of `Qiskit-Aer`, the classical simulation framework of IBM's quantum programming suite `Qiskit` [4]. This enables classical simulation and verification of quantum circuits built in Qiskit on system sizes above 30 qubits, which quickly become inaccessible with the default vector-based method. This version also supports parallel processing over the stabilizer state decomposition, which improves the performance of the Metropolis step.

Let us briefly comment on how simulators based on the stabilizer rank compare with quasi-probability methods [20, 37, 47]. The latter use a discrete Wigner function representation of quantum states and Monte Carlo sampling to approximate a given output probability of the target circuit with a small additive error. Negativity of the Wigner function is an important parameter that quantifies severity of the "sign problem" associated with the Monte Carlo sampling. The negativity also controls the runtime of quasi-probability methods. For example, the simulator proposed in [47] has runtime $\epsilon^{-2}M^2$, where M is the negativity and ϵ is the desired approximation error. In contrast to stabilizer rank simulators, quasi-probability methods do not directly apply to stabilizer operations on qubits since the latter are not known to have a non-negative Wigner function representation [20, 36]. Furthermore, such methods are not well-suited for sampling the output distribution since this task requires a small *multiplicative* error in approximating individual output probabilities.

Our work leaves several open questions. Since the efficiency of Clifford simulators hinges on the ability to find low-rank stabilizer decompositions of multi-qubit magic states, improved techniques for finding such decompositions are of great interest. For example, consider a magic state $|\psi\rangle = U|+\rangle^{\otimes n}$, where U is a diagonal circuit composed of Z, CZ, and CCZ gates. We anticipate that a low-rank exact stabilizer decomposition of ψ can be found by computing the *transversal number* [5] of a suitable hypergraph describing the placement of CCZ gates. Such low-rank decompositions may lead to more efficient simulation algorithms for Clifford+CCZ circuits. We leave as an open question whether the stabilizer extent $\xi(\psi)$ is multiplicative under tensor products for general states ψ. Finally, it is of great interest to derive lower bounds on the stabilizer rank of n-qubit magic states scaling exponentially with n.

4 Subroutines

Throughout this section we use the following notations. Suppose $x \in \{0,1\}^n$ is a bit string. We shall consider x as a row vector and write x^T for the transposed column vector. The Hamming weight of x denoted $|x|$ is the number of ones in x. The support of x is the subset of indices $j \in [n]$ such that $x_j = 1$. Given a single-qubit operator P let $P(x)$ be an n-qubit product operator that applies P to each qubit in the support of x, that is, $P(x) = P^{x_1} \otimes \cdots \otimes P^{x_n}$. We shall use the notation $\oplus$ for the addition of binary vectors modulo two. Let $x \cdot y \equiv \sum_{j=1}^{n} x_j y_j$.

4.1 Phase-sensitive Clifford simulator

In this section we describe a Clifford simulator based on stabilizer tableau [2] that keeps track of the global phase of stabilizer states. We shall consider Clifford circuits expressed using a gate set

$$S, \quad CZ, \quad CX, \quad H. \tag{40}$$

Here CZ and CX are controlled-Z and -X gates, H is the Hadamard gate, and $S = |0\rangle\langle 0| + i|1\rangle\langle 1|$.

First let us define a data format to describe stabilizer states. Suppose U is a unitary Clifford operator. We say that U is a control-type or *C-type* operator if

$$U|0^n\rangle = |0^n\rangle. \tag{41}$$

For example, the gates S, CZ, CX and any product of such gates are C-type operators. We say that U is a Hadamard-type or *H-type* operator if U is a tensor product of the Hadamard and the identity gates. Previously known results on canonical decompositions of Clifford circuits [25, 42, 54] imply that any n-qubit stabilizer state ϕ can be expressed as

$$|\phi\rangle = \omega U_C U_H |s\rangle, \tag{42}$$

where U_C and U_H are C-type and H-type Clifford operators, $s \in \{0,1\}^n$ is a basis vector, and ω is a complex number. We shall refer to the decomposition Eq. (42) as a CH-form of ϕ. Note that this form may be non-unique.

We shall describe the unitary U_C by its stabilizer tableaux, that is, a list of Pauli operators $U_C^{-1} Z_p U_C$ and $U_C^{-1} X_p U_C$. The global phase of U_C is fixed by Eq. (41). Using Eq. (41) one can check that $U_C^{-1} Z_p U_C$ is a tensor product of Pauli Z and the identity operators I. Thus the stabilizer tableaux of U_C can be described by binary matrices F, G, M of size $n \times n$ and a phase vector $\gamma \in \mathbb{Z}_4^n$ such that

$$U_C^{-1} Z_p U_C = \prod_{j=1}^n Z_j^{G_{p,j}} \quad \text{and} \quad U_C^{-1} X_p U_C = i^{\gamma_p} \prod_{j=1}^n X_j^{F_{p,j}} Z_j^{M_{p,j}} \tag{43}$$

for all $p = 1, \dots, n$. Here $X^0 \equiv Z^0 \equiv I$. We shall describe the unitary U_H by a string $v \in \{0,1\}^n$ such that

$$U_H = H(v) \equiv H_1^{v_1} \otimes H_2^{v_2} \otimes \cdots \otimes H_n^{v_n}. \tag{44}$$

To summarize, the CH-form is fully specified by the data $(F, G, M, \gamma, v, s, \omega)$. Let us agree that $\omega = 1$ whenever it is omitted.

Below we describe an algorithm that takes as input a sequence of Clifford gates $U_1, \dots, U_m$ from the gate set Eq. (40) and outputs the CH-form of a stabilizer state

$$|\phi\rangle = U_m \cdots U_2 U_1 |0^n\rangle. \tag{45}$$

The runtime is $O(n)$ per each gate S, CZ, CX and $O(n^2)$ per each Hadamard gate. We also show how to compute an amplitude $\langle x|\phi\rangle$ and sample x from the distribution $|\langle x|\phi\rangle|^2$ assuming that ϕ is specified by its CH-form. These tasks take time $O(n^2)$. Finally, we consider projective gates $(I+P)/2$, where P is a Pauli operator. We show how to simulate projective gates in time $O(n^2)$.

Simulation of unitary gates. The initial state $|0^n\rangle$ has a trivial CH-form with $s = 0^n$ and $U_C = U_H = I$. Thus we initialize the CH data as $G = F = I$, M is the zero matrix, and γ, v, s are zero vectors. Suppose ϕ is a stabilizer state with the CH form

$$|\phi\rangle = U_C U_H |s\rangle$$

described by the data (F, G, M, γ, v, s). Consider a gate $\Gamma \in \{S, CZ, CX, H\}$ applied to some subset of qubits. The state $\Gamma|\phi\rangle$ has a CH-form

$$\Gamma|\phi\rangle = \Gamma U_C U_H |s\rangle = \omega' U_C' U_H' |s'\rangle \tag{46}$$

with the corresponding data $(F', G', M', \gamma', v', s', \omega')$. Let us show how to compute this data.

The case $\Gamma \in \{S, CZ, CX\}$ is trivial: one can absorb Γ into the C-layer obtaining $U_C' = \Gamma U_C$. The stabilizer tableaux of U_C is updated using the standard Aaronson-Gottesman algorithm [2] (explicit update rules are provided at the end of this section). This update takes time $O(n)$.

Let $\Gamma = H_p$ be the Hadamard gate applied to a qubit $p \in [n]$. Commuting H_p through the C- and H-layer using the identity $H_p = 2^{-1/2}(X_p + Z_p)$ and Eq. (43) one gets

$$H_p|\phi\rangle = 2^{-1/2} U_C U_H [(-1)^\alpha |t\rangle + i^{\gamma_p} (-1)^\beta |u\rangle), \tag{47}$$

where $t, u \in \{0,1\}^n$ are defined by

$$t_j = s_j \oplus G_{p,j} v_j \quad \text{and} \quad u_j = s_j \oplus F_{p,j} \bar{v}_j \oplus M_{p,j} v_j \tag{48}$$

for $j = 1, \dots, n$. Here and below $\bar{v}_j \equiv 1 - v_j$. Furthermore,

$$\alpha = \sum_{j=1}^n G_{p,j} \bar{v}_j s_j \quad \text{and} \quad \beta = \sum_{j=1}^n M_{p,j} \bar{v}_j s_j + F_{p,j} v_j (M_{p,j} + s_j). \tag{49}$$

The case $t = u$ is trivial: Eq. (47) gives the desired CH-form of $H_p|\phi\rangle$ with $s' = t = u$ and $\omega' = 2^{-1/2}[(-1)^\alpha + i^{\gamma_p}(-1)^\beta]$. From now on assume that $t \neq u$.

Proposition 4. *Suppose $t, u \in \{0,1\}^n$ are distinct strings and $\delta \in \mathbb{Z}_4$. Then the state $U_H(|t\rangle + i^\delta |u\rangle)$ has a CH-form*

$$U_H(|t\rangle + i^\delta |u\rangle) = \omega W_C W_H |s'\rangle, \tag{50}$$

where the C-layer W_C consists of $O(n)$ gates from the set $\{S, CZ, CX\}$. The decomposition Eq. (50) can be computed in time $O(n)$.

Choosing $\delta = \gamma_p + 2(\alpha + \beta) \pmod 4$ and substituting Eq. (50) into Eq. (47) one gets

$$H_p|\phi\rangle = 2^{-1/2}(-1)^\alpha \omega \cdot U_C W_C \cdot W_H |s'\rangle. \tag{51}$$

This gives the desired CH-form of $H_p|\phi\rangle$ with

$$\omega' = 2^{-1/2}(-1)^\alpha \omega, \quad U'_C = U_C W_C, \quad U'_H = W_H. \tag{52}$$

Finally, one needs to compute the stabilizer tableaux of U'_C. Since W_C consists of $O(n)$ gates S, CZ, CX it suffices to give update rules for the stabilizer tableaux of U_C under the right multiplications $U_C \leftarrow U_C \Gamma$ with $\Gamma \in \{S, CZ, CX\}$. Explicit update rules are provided at the end of this section (this is a straightforward application of the stabilizer formalism). Each update rule takes time $O(n)$. Since W_C contains $O(n)$ gates, the full simulation cost of the Hadamard gate is $O(n^2)$.

Proof of Prosposition 4. We shall construct a C-type circuit V_C and bit strings $y, z \in \{0,1\}^n$ such that

- y and z differ on a single bit $q \in [n]$,
- $U_H|t\rangle = V_C U_H |y\rangle$,
- $U_H|u\rangle = V_C U_H |z\rangle$.

Then

$$U_H(|t\rangle + i^\delta |u\rangle) = V_C U_H(|y\rangle + i^\delta |z\rangle). \tag{53}$$

Since $y_i = z_i$ for $i \neq q$ and $y_q \neq z_q$, the state $U_H(|y\rangle + i^\delta |z\rangle)$ is a tensor product of single-qubit states $H^{v_i}|y_i\rangle$ on qubits $i \neq q$ and a stabilizer state $H^{v_q}(|y_q\rangle + i^\delta |z_q\rangle)$ on qubit q. Let us write

$$H^{v_q}(|y_q\rangle + i^\delta |z_q\rangle) = \omega S^a H^b |c\rangle$$

for some $a, b, c \in \{0,1\}$ and some complex number ω. We arrive at

$$U_H(|t\rangle + i^\delta |u\rangle) = \omega (V_C S_q^a)(U_H H_q^{b \oplus v_q})|s'\rangle,$$

where $s'_q = c$ and $s'_i = y_i = z_i$ for $i \neq q$. This is the desired form Eq. (50) with $W_C = V_C S_q^a$ and $W_H = U_H H_q^{b \oplus v_q}$.

It remains to construct V_C, y, z as above. We shall choose V_C such that

$$U_H V_C U_H = \prod_{i \in [n] \setminus q \,:\, t_i \neq u_i} CX_{q,i} \tag{54}$$

for some qubit $q \in [n]$ such that $t_q \neq u_q$. The circuit in the righthand side of Eq. (54) maps t, u to strings y, z that differ only on the q-th bit. Accordingly, $V_C U_H |t\rangle = U_H |y\rangle$ and $V_C U_H |u\rangle = U_H |z\rangle$, as desired.

For each $b \in \{0,1\}$ define a subset

$$\mathcal{V}_b = \{i \in [n] \,:\, v_i = b \quad \text{and} \quad t_i \neq u_i\}.$$

Here $v \in \{0,1\}^n$ defines the H-layer U_H, see Eq. (44). By assumption, at least one of the subsets $\mathcal{V}_b$ is non-empty.

Suppose first that $\mathcal{V}_0 \neq \emptyset$. Let q be the first qubit of $\mathcal{V}_0$. Define

$$V_C = \prod_{i \in \mathcal{V}_0 \setminus q} CX_{q,i} \cdot \prod_{i \in \mathcal{V}_1} CZ_{q,i}.$$

Here $CX_{q,i}$ has control q and target i. If $\mathcal{V}_0 = \{q\}$ then gates $CX_{q,i}$ are skipped. Likewise, if $\mathcal{V}_1 = \emptyset$ then the gates $CZ_{q,i}$ are skipped. Simple algebra shows that V_C obeys Eq. (54).

Suppose now that $\mathcal{V}_0 = \emptyset$. Then $\mathcal{V}_1 \neq \emptyset$ since $t \neq u$. Let q be the first qubit of $\mathcal{V}_1$. Define

$$V_C = \prod_{i \in \mathcal{V}_1 \setminus q} CX_{i,q}.$$

Let us agree that $V_C = I$ if $\mathcal{V}_1 = \{q\}$. Simple algebra shows that V_C obeys Eq. (54).

In both cases the strings y, z have the form

$$\text{if } t_q = 1 \text{ then } y = u \oplus e_q \text{ and } z = u,$$

$$\text{if } t_q = 0 \text{ then } y = t \text{ and } z = t \oplus e_q.$$

Here $e_q \in \{0,1\}^n$ is a string with a single non-zero at the q-th bit. □

In the rest of this section we provide rules for updating the stabilizer tableaux of U_C under the left and the right multiplications $U_C \leftarrow \Gamma U_C$ and $U_C \leftarrow U_C \Gamma$, where Γ is one of the gates S, CZ, CX. We shall write $\mathcal{L}[\Gamma]$ and $\mathcal{R}[\Gamma]$ for the left and the right multiplication by Γ. Below $p = 1, \ldots, n$. All phase vector updates are performed modulo four.

$$\mathcal{R}[S_q] : \begin{cases} M_{p,q} & \leftarrow & M_{p,q} \oplus F_{p,q} \\ \gamma_p & \leftarrow & \gamma_p - F_{p,q} \end{cases} \qquad \mathcal{L}[S_q] : \begin{cases} M_{q,p} & \leftarrow & M_{q,p} \oplus G_{q,p} \\ \gamma_q & \leftarrow & \gamma_q - 1 \end{cases}$$

$$\mathcal{R}[CZ_{q,r}] : \begin{cases} M_{p,q} & \leftarrow & M_{p,q} \oplus F_{p,r} \\ M_{p,r} & \leftarrow & M_{p,r} \oplus F_{p,q} \\ \gamma_p & \leftarrow & \gamma_p + 2F_{p,q}F_{p,r} \end{cases} \qquad \mathcal{L}[CZ_{q,r}] : \begin{cases} M_{q,p} & \leftarrow & M_{q,p} \oplus G_{r,p} \\ M_{r,p} & \leftarrow & M_{r,p} \oplus G_{q,p} \end{cases}$$

$$\mathcal{R}[CX_{q,r}] : \begin{cases} G_{p,q} & \leftarrow & G_{p,q} \oplus G_{p,r} \\ F_{p,r} & \leftarrow & F_{p,r} \oplus F_{p,q} \\ M_{p,q} & \leftarrow & M_{p,q} \oplus M_{p,r} \end{cases} \qquad \mathcal{L}[CX_{q,r}] : \begin{cases} G_{r,p} & \leftarrow & G_{r,p} \oplus G_{q,p} \\ F_{q,p} & \leftarrow & F_{q,p} \oplus F_{r,p} \\ M_{q,p} & \leftarrow & M_{q,p} \oplus M_{r,p} \\ \gamma_q & \leftarrow & \gamma_q + \gamma_r + 2(MF^T)_{q,r} \end{cases}$$

Simulating measurements. Let $x \in \{0,1\}^n$ be a basis vector. Using Eqs. (41,43) one gets

$$\langle x|U_C U_H|s\rangle = \langle 0^n| \left(\prod_{p=1}^{n} U_C^{-1} X_p^{x_p} U_C \right) U_H|s\rangle \equiv \langle 0^n|Q U_H|s\rangle. \tag{55}$$

Note that Q is a product of $|x|$ Pauli operators that appear in Eq. (43). It can be computed inductively in time $O(n^2)$ by setting $Q = I$ and performing updates $Q \leftarrow Q \cdot U_C^{-1} X_p^{x_p} U_C$ for each p with $x_p = 1$. Write $Q = i^\mu Z(t)X(u)$ for some $\mu \in \mathbb{Z}_4$ and $t, u \in \{0,1\}^n$. Note that $u = xF \pmod 2$. Then

$$\langle x|U_C U_H|s\rangle = \langle 0^n|Q U_H|s\rangle = 2^{-|v|/2} i^\mu \prod_{j\,:\,v_j=1} (-1)^{u_j s_j} \prod_{j\,:\,v_j=0} \langle u_j|s_j\rangle. \tag{56}$$

Thus computing the amplitude $\langle x|U_C U_H|s\rangle$ takes time $O(n^2)$.

Consider a probability distribution $P(x) = |\langle x|U_C U_H|s\rangle|^2$. From Eq. (56) one infers that $P(x) = 2^{-|v|}$ if $u_j = s_j$ for all bits j with $v_j = 0$ and $P(x) = 0$ otherwise. Since U_C preserves the Pauli commutation rules, one has $FG^T = I \pmod 2$. Thus $x = wG^T \pmod 2$, where $w \in \{0,1\}^n$ is a row vector satisfying $w_j = s_j$ if $v_j = 0$. The remaining bits of w are picked uniformly at random. Thus one can sample x from $P(x)$ as follows:

- Set $w = s$.
- For each j such that $v_j = 1$ flip the j-th bit of w with probability $1/2$.
- Output $x = wG^T \pmod 2$.

This takes time $O(n^2)$. Finally, consider a projective gate $\Gamma = (I + P)/2$, where $P = P^\dagger$ is a Pauli operator. We have

$$\Gamma|\phi\rangle = \Gamma U_C U_H|s\rangle = (1/2) U_C U_H (I + Q)|s\rangle,$$

where Q is a Pauli operator that can be computed in time $O(n^2)$ using the stabilizer tableaux of U_C. Write $(I + Q)|s\rangle = |s\rangle + i^\delta |t\rangle$ for some $t \in \{0,1\}^n$ and $\delta \in \mathbb{Z}_4$. We can now compute the CH-form of $\Gamma|\phi\rangle$ using Proposition 4 in the same fashion as was done above for the Hadamard gate.

4.2 Heuristic Metropolis simulator

Consider a state $|\psi\rangle = \sum_{\alpha=1}^{k} b_\alpha |\phi_\alpha\rangle$, where $\phi_1, \ldots, \phi_k$ are n-qubit stabilizer states. We assume that all states ϕ_α are specified by their CH-form. This form can be efficiently computed using the Clifford simulator of Section 4.1. Our goal is to sample $x \in \{0,1\}^n$ from the probability distribution

$$P(x) = \frac{|\langle x|\psi\rangle|^2}{\|\psi\|^2}.$$

To this end define a Metropolis-type Markov chain $\mathcal{M}$ with a state space $\Omega = \{x \in \{0,1\}^n : P(x) > 0\}$. Suppose the current state of the chain $x \in \Omega$. Then the next state x' is generated as follows.

- Pick an integer $j \in [n]$ uniformly at random and let $y = x \oplus e_j$.
- If $P(y) \geq P(x)$ then set $x' = y$.
- Otherwise generate a random bit $b \in \{0,1\}$ such that $\Pr(b = 1) = P(y)/P(x)$.
- If $b = 1$ then set $x' = y$. Otherwise set $x' = x$.

We shall refer to the mapping $x \to x'$ as a Metropolis step. Let us make a simplifying assumption that the chain $\mathcal{M}$ is irreducible, that is, for any pair of strings $x, y \in \Omega$ there exist a path $x^0 = x, x^1, \ldots, x^L = y \in \Omega$ such that x^i and x^{i+1} differ on a single bit for all i. Then $P(x)$ is the unique steady distribution of $\mathcal{M}$. One can (approximately) sample x from $P(x)$ by implementing $T \gg 1$ Metropolis steps starting from some (random) initial state $x_{in} \in \Omega$ and using the final state as the output string.

We claim that one can implement T Metropolis steps in time

$$O(knT) + O(kn^2).$$

Here the term $O(kn^2)$ is the cost of computing the initial probability $P(x_{in})$ using the algorithm of Section 4.1. Indeed, suppose we have already implemented several steps reaching some state $x \in \Omega$. Let $y = x \oplus e_j$ be a proposed next state. Consider some fixed stabilizer state $\phi \equiv \phi_\alpha$ that contributes to ψ and let $|\phi\rangle = U_C U_H |s\rangle$ be its CH-form. Then

$$\langle y|\phi\rangle = \langle x \oplus e_j | U_C U_H |s\rangle = \langle 0^n | U_C^{-1} X_j U_C \cdot Q_x U_H |s\rangle, \tag{57}$$

where

$$Q_x \equiv \prod_{p=1}^{n} U_C^{-1} X_p^{x_p} U_C.$$

Note that computing Q_x for the initial state $x = x_{in}$ and $\alpha = 1, \ldots, k$ takes time $O(kn^2)$. Suppose Q_x has been already computed. Since $U_C^{-1} X_j U_C$ is determined by the stabilizer tableaux of U_C, see Eq. (43), one can compute the product $Q_y = U_C^{-1} X_j U_C \cdot Q_x$ in time $O(n)$. Then the amplitude $\langle y|\phi\rangle = \langle 0^n | Q_y U_H |s\rangle$ can be computed in time $O(n)$. This shows that the ratio

$$\frac{P(y)}{P(x)} = \left| \frac{\sum_{\alpha=1}^{k} b_\alpha \langle y|\phi_\alpha\rangle}{\sum_{\alpha=1}^{k} b_\alpha \langle x|\phi_\alpha\rangle} \right|^2.$$

can be computed in time $O(kn)$ provided that one saves the Pauli Q_x for each stabilizer term ϕ_α after each Metropolis step. This achieves the runtime scaling quoted above.

In general there is no reason to expect that the Metropolis chain defined above is irreducible. Furthermore, its mixing time is generally unknown. Thus the proposed algorithm should be considered as a heuristic. However, the numeric results shown in Fig. **??** were obtained using the Metropolis method to sample from the output distribution of the QAOA circuit.

We expect that the Metropolis chain may be rapidly mixing in the case when ψ approximates the output state of some small-depth quantum circuit. In particular, if $P(x)$ is the exact output distribution of a constant-depth circuit and each Metropolis step flips $O(1)$ bits, one can use isoperimetric inequalities derived in Refs. [18, 21] to show that $P(x)$ is the uniqiue steady state of $\mathcal{M}$ and its mixing time is at most $poly(n)$.

4.3 Fast norm estimation

As before, consider a state $|\psi\rangle = \sum_{\alpha=1}^{k} b_\alpha |\phi_\alpha\rangle$, where $\phi_1, \ldots, \phi_k$ are n-qubit stabilizer states specified by their CH-form. Recall that our goal is to estimate the norm $\|\psi\|^2$ and to sample the probability distribution $P(x) \sim |\langle x|\psi\rangle|^2$. In this section we describe an algorithm that takes as input the target state ψ, error tolerance parameters $\epsilon, \delta > 0$, and outputs a random number η such that

$$(1-\epsilon)\|\psi\|^2 \le \eta \le (1+\epsilon)\|\psi\|^2 \tag{58}$$

with probability at least $1-\delta$. The algorithm has runtime

$$O(kn^3\epsilon^{-2}\log\delta^{-1}). \tag{59}$$

The key idea proposed in Ref. [11] is to estimate $\|\psi\|^2$ by computing inner products between ψ and randomly chosen stabilizer states ϕ. It can be shown [11] that the quantity $\eta \equiv 2^n|\langle\phi|\psi\rangle|^2$ is an unbiased estimator of $\|\psi\|^2$ with the standard deviation $\approx \|\psi\|^2$, provided that ϕ is drawn from the uniform distribution on the set of stabilizer states. Thus the empirical mean of η provides an estimate of $\|\psi\|^2$ with a small multiplicative error. The quantity η can be computed in time $O(kn^3)$ since $\langle\phi|\psi\rangle = \sum_{\alpha=1}^{k} b_\alpha \langle\phi|\phi_\alpha\rangle$ and the inner product between stabilizer states can be computed in time $O(n^3)$.

Here we improve upon the algorithm of Ref. [11] in two respects. First, we show that the random stabilizer state ϕ used in the norm estimation method can be drawn from a certain subset of stabilizer states that we call equatorial states. By definition, a stabilizer state ϕ is called equatorial iff it has equal amplitude on each basis vector. Sampling an equatorial state from the uniform distribution is particularly simple: all it takes is tossing an unbiased coin $O(n^2)$ times. Secondly, we greatly simplify computation of the inner products $\langle\phi|\phi_\alpha\rangle$. This is achieved by using the CH-form to describe stabilizer states and by introducing a more efficient (and simpler) algorithm for computing certain exponential sums (see Lemma 4 below).

We shall now formally describe the norm estimation algorithm. Let $\mathcal{M}_n$ be the set of symmetric $n \times n$ matrices M with off-diagonal entries $\in \{0,1\}$ and diagonal entries $\in \{0,1,2,3\}$. For any matrix $A \in \mathcal{M}_n$ define a stabilizer state

$$|\phi_A\rangle = 2^{-n/2} \sum_{x\in\{0,1\}^n} i^{xAx^T} |x\rangle. \tag{60}$$

We shall refer to ϕ_A as an *equatorial state* (note that ϕ_A lies on the equator of the Bloch sphere for $n=1$).

Lemma 2 (**Norm Estimation**). *Let ψ be an arbitrary n-qubit state. Define a random variable*

$$\eta_A = 2^n|\langle\phi_A|\psi\rangle|^2, \tag{61}$$

where $A \in \mathcal{M}_n$ is chosen uniformly at random. Then η_A has mean $\|\psi\|^2$ and its variance is at most $\|\psi\|^4$.

Lemma 3 (**Inner Product**). *Suppose $|\phi\rangle = U_C U_H |s\rangle$ is a stabilizer state in the CH-form, where $U_H = H(v)$ and U_C has a stabilizer tableaux (F,G,M,γ). Suppose ϕ_A is an equatorial state specified by a matrix $A \in \mathcal{M}_n$. Define a matrix $J \in \mathcal{M}_n$ such that $\mathrm{diag}(J) = \gamma$ and $J_{a,b} = (MF^T)_{a,b} \pmod 2$ for $a \ne b$. Define a matrix*

$$K = G^T(A+J)G.$$

Then

$$\langle\phi|\phi_A\rangle = 2^{-(n+|v|)/2} \cdot i^{sKs^T} \cdot (-1)^{s\cdot v} \sum_{x\le v} i^{xKx^T + 2x(s+sK)^T}. \tag{62}$$

Here the sum is over n-bit strings x satisfying $x_j \le v_j$ for all j.

Since the Pauli operators $U_C^{-1} X_p U_C$ pairwise commute, $MF^T \pmod 2$ is a symmetric matrix, see Eq. (43). Therefore K is a symmetric matrix and thus i^{xKx^T} depends only on off-diagonal

elements of K modulo two and diagonal elements of K modulo four. Thus the sum that appears in Eq. (62) can be expressed as

$$\mathcal{Z}(B) = \sum_{x\in\{0,1\}^{|v|}} i^{xBx^T}$$

for a suitable matrix $B \in \mathcal{M}_{|v|}$, namely, a restriction of the matrix $K + 2\mathrm{diag}(s + sK)$ onto the subset of rows and columns j with $v_j = 1$. We shall refer to $\mathcal{Z}(B)$ as an exponential sum associated with B.

Lemma 4 (Exponential Sum). *There is a deterministic algorithm with a runtime $O(n^3)$ that takes as input a matrix $B \in \mathcal{M}_n$ and outputs integers $p, q \geq 0$ and $\alpha, \beta \in \{0, 1\}$ such that $\mathcal{Z}(B) = \alpha 2^p + i\beta 2^q$.*

The desired estimate of $\|\psi\|^2$ can now be obtained by sampling i.i.d. random matrices $A_1, \ldots, A_L \in \mathcal{M}_n$ and computing the empirical mean $\eta = L^{-1}(\eta_{A_1} + \ldots + \eta_{A_L})$. Indeed, Lemma 2 and the Chebyshev inequality imply that η achieves the desired approximation Eq. (58) with probability at least $3/4$ if $L = 4\epsilon^{-2}$. The error probability can be reduced to any desired level δ by generating $K = O(\log \delta^{-1})$ independent estimates $\eta^1, \ldots, \eta^K$ as above such that each estimate η^a satisfies Eq. (58) with probability at least $3/4$. Let η_{med} be the median of $\eta^1, \ldots, \eta^K$. Then standard arguments show that η_{med} satisfies Eq. (58) with probability at least $1-\delta$. Computing each sample η_{A_i} using Lemmas 3,4 takes time $O(kn^3)$. Since the total number of samples is $KL = O(\epsilon^{-2} \log \delta^{-1})$, we arrive at Eq. (59).

Finally, let us sketch how to use the norm estimation for sampling $x \in \{0, 1\}^n$ from a distribution

$$P(x) = \frac{|\langle x|\psi\rangle|^2}{\|\psi\|^2}.$$

Let $P_w(x_1, \ldots, x_w)$ be the marginal distribution describing the first w bits. We have $P_w(x) = \|\Pi\psi\|^2/\|\psi\|^2$, where Π projects the j-th qubit onto the state x_j for $1 \leq j \leq w$. It can be written as

$$\Pi = 2^{-w} \prod_{j=1}^{w} (I + (-1)^{x_j} Z_j)$$

One can compute a rank-k stabilizer decomposition of the state $\Pi|\psi\rangle$ in time $O(kwn^2)$ using the Clifford simulator of Section 4.1. By estimating the norms $\|\psi\|^2$ and $\|\Pi\psi\|^2$ one can approximate any marginal probability $P_w(x)$ with a small multiplicative error. In the same fashion one can approximate conditional probabilities

$$P_w(x_w|x_1, \ldots, x_{w-1}) = \frac{P_w(x_1, \ldots, x_w)}{P_{w-1}(x_1, \ldots, x_{w-1})}.$$

Now one can sample the bits of x one by one using the chain rule

$$P(x) = P_1(x_1)P_2(x_2|x_1) \cdots P_n(x_n|x_1, \ldots, x_{n-1}).$$

Clearly, the same method can be used to sample any marginal distribution of $P(x)$.

To avoid accumulation of errors, each of $O(n)$ steps in the chain rule requires an estimate of the marginal probabilities $P_w(x)$ with a multiplicative error $O(n^{-1})$. (This guarantees that the full probability $P(x)$ is estimated using the chain rule within a small multiplicative error.) This would require setting the precision ϵ in the norm estimation method as $\epsilon = O(n^{-1})$. Thus the cost of each norm estimation would be $O(kn^3\epsilon^{-2}) = O(kn^5)$. Since the total number of norm estimations is $\Omega(n)$, the overall runtime for generating a single sample from $P(x)$ with a small error would scale as $O(kn^6)$. This quickly becomes impractical. However, if our goal is to sample only w bits from $P(x)$, a similar analysis shows that the overall runtime scales as $O(kn^3w^3)$. Thus the sampling method based on the norm estimation is practical only for small values of w. In contrast, Metropolis simulator allows one to sample all n output bits and has runtime $O(knT)$, where T is the mixing time (which is generally unknown).

In the rest of this section we prove Lemmas 2,3,4.

Proof of Lemma 2. Let

$$Q_1 = \mathbb{E}_A |\phi_A\rangle\langle\phi_A| \quad \text{and} \quad Q_2 = \mathbb{E}_A |\phi_A\rangle\langle\phi_A|^{\otimes 2}.$$

Since the distribution of A is invariant under shifts $A_{j,j} \leftarrow A_{j,j}+2$, one concludes that Q_1 commutes with single-qubit Pauli-Z operators. Thus Q_1 is diagonal in the Z-basis. Furthermore, all diagonal matrix elements of $|\phi_A\rangle\langle\phi_A|$ are equal to 2^{-n}. This proves $Q_1 = 2^{-n} I$ and thus η_A has expected value $2^n \langle\psi|Q_1|\psi\rangle = \|\psi\|^2$.

By definition,

$$Q_2 = 4^{-n} \sum_{w,x,y,z} E(w,x,y,z) \cdot |w,x\rangle\langle y,z| \quad \text{where} \quad E(w,x,y,z) = \mathbb{E}_A \, i^{wAw^T + xAx^T - yAy^T - zAz^T}.$$

Here the sum runs over all n-bit strings. We shall use the following fact.

Proposition 5 (Ref. [15]). *$E(w,x,y,z) = 0$ unless $w + x = y + z \pmod 4$ and at least two of the strings w, x, y coincide.*

Proof. By definition, diagonal entries $A_{p,p} \in \mathbb{Z}_4$ and off-diagonal entries $A_{p,q} = A_{q,p} \in \mathbb{Z}_2$ are i.i.d. uniform random variables. The entry $A_{p,p}$ contributes a factor $i^{A_{p,p}(w_p + x_p - y_p - z_p)}$ to $E(w,x,y,z)$. Thus $E(w,x,y,z) = 0$ unless

$$w_p + x_p = y_p + z_p \pmod 4 \tag{63}$$

for all p. This proves the first claim. The entry $A_{p,q} = A_{q,p}$ contributes a factor

$$(-1)^{A_{p,q}(w_p w_q + x_p x_q - y_p y_q - z_p z_q)}$$

to $E(w,x,y,z)$. Thus $E(w,x,y,z) = 0$ unless

$$w_p w_q + x_p x_q - y_p y_q - z_p z_q = 0 \pmod 2. \tag{64}$$

From Eq. (63) one gets $z_p = w_p + x_p + y_p \pmod 2$. Substituting this expression for z_p into Eq. (64) one concludes that $E(w,x,y,z) = 0$ unless

$$(w_p x_q + w_q x_p) + (x_p y_q + x_q y_p) + (y_p w_q + y_q w_p) = 0 \pmod 2 \tag{65}$$

for all $p < q$. If $w = x = y$ then there remains nothing to prove. Otherwise, there exists an index $p \in [n]$ such that exactly two of the variables w_p, x_p, y_p coincide. Since Eq. (65) is symmetric under permutations of w, x, y, assume wlog that $x_p = y_p \neq w_p$. Consider two cases.
Case 1: $x_p = y_p = 0$ and $w_p = 1$. Substituting this into Eq. (65) one gets $y_q = x_q$ for all $q \neq p$. Thus $x = y$.
Case 2: $x_p = y_p = 1$ and $w_p = 0$. Substituting this into Eq. (65) one gets $y_q + x_q + w_q + w_q = 0 \pmod 2$ for all $q \neq p$, that is, $x = y$.
We conclude that at least two of the strings w, x, y coincide. □

Let us consider the cases when $E(w,x,y,z) \neq 0$. *Case 1:* $w = x$. Then $y + z = 2x \pmod 4$ which is possible only if $y = z$ and thus $w = x = y = z$. *Case 2:* $w = y$. Then $x = z$ and $E(y,x,y,x) = 1$. *Case 3:* $w = z$. Then $x = y$ and $E(z,x,x,z) = 1$. The above shows that non-zero contributions to Q_2 come only from the terms $E(w,x,w,x) = E(w,x,x,w) = 1$. Thus

$$Q_2 = 4^{-n}(I + \mathrm{SWAP}) - 4^n \sum_x |x,x\rangle\langle x,x|,$$

Here the last term is introduced to avoid overcounting since the terms with $w = x = y = z$ appear in all three cases. We arrive at

$$\mathbb{E}_A(\eta_A^2) = 4^n \langle\psi^{\otimes 2}|Q_2|\psi^{\otimes 2}\rangle \leq \langle\psi^{\otimes 2}|I + \mathrm{SWAP}|\psi^{\otimes 2}\rangle = 2\|\psi\|^4.$$

It follows that η_A has variance at most $\|\psi\|^4$. □

Proof of Lemma 3. Let $a \in \{0,1\}^n$ be an arbitrary string. From Eq. (43) one easily gets

$$U_C^{-1} X(a) U_C = \prod_{p=1}^{n} U_C^{-1} X_p^{a_p} U_C = i^{aJa^T} \cdot X(aF \pmod 2) Z(aM \pmod 2).$$

Here $J \in \mathcal{M}_n$ is defined in the statement of the lemma. It follows that

$$U_C^{-1}|a\rangle = U_C^{-1} X(a) U_C |0^n\rangle = i^{aJa^T} |aF \pmod 2\rangle.$$

Therefore

$$U_C^{-1}|\phi_A\rangle = 2^{-n/2} \sum_{x\in\{0,1\}^n} i^{x(A+J)x^T} |xF \pmod 2\rangle.$$

Recall that $FG^T \pmod 2 = I$. Perform a change of variable $x = yG^T \pmod 2$. Then $x = yG^T + 2u$ for some integer vector u. Using the fact that A and J are symmetric matrices one gets

$$x(A+J)x^T = yG^T(A+J)Gy^T + 4u(A+J)Gy^T + 4u(A+J)u^T.$$

Denoting $K = G^T(A+J)G$ one gets

$$U_C^{-1}|\phi_A\rangle = 2^{-n/2} \sum_{y\in\{0,1\}^n} i^{yKy^T} |y\rangle. \tag{66}$$

We have

$$U_H|s\rangle = 2^{-|v|/2} \sum_{x\le v} (-1)^{s\cdot v + s\cdot x} |s\oplus x\rangle, \tag{67}$$

Taking the inner product of the states Eqs. (66,67) gives

$$\langle\phi|\phi_A\rangle = \langle s|U_H U_C^{-1}|\phi_A\rangle = 2^{-(n+|v|)/2}(-1)^{s\cdot v} \sum_{x\le v} (-1)^{s\cdot x} \cdot i^{(s\oplus x)K(s\oplus x)^T}. \tag{68}$$

Writing $s\oplus x = s + x + 2u$ for some integer vector u and using the fact that K is symmetric one gets

$$(s\oplus x)K(s\oplus x)^T = (s+x)K(s+x)^T + 4uK(s+x)^T + 4uKu^T.$$

It follows that

$$i^{(s\oplus x)K(s\oplus x)^T} = i^{sKs^T + xKx^T + 2xKs^T}.$$

Combining this and Eq. (68) proves Eq. (62). □

Proof of Lemma 4. Define a binary upper-triangular matrix M of size $n\times n$ such that $M_{\alpha,\beta} = B_{\alpha,\beta}$ for $\alpha < \beta$. Define binary vectors $L, K \in \{0,1\}^n$ such that $B_{\alpha,\alpha} = 2L_\alpha + K_\alpha$ for all α. Then $i^{xBx^T} = i^{q(x)}$, where $q : \{0,1\}^n \to \mathbb{Z}_4$ is a binary quadratic form defined as

$$q(x) = 2 \sum_{1\le\alpha<\beta\le n} M_{\alpha,\beta} x_\alpha x_\beta + \sum_{1\le\alpha\le n} (2L_\alpha + K_\alpha) x_\alpha \pmod 4. \tag{69}$$

Our goal is to compute the exponential sum

$$\mathcal{Z} \equiv \sum_{x\in\{0,1\}^n} i^{q(x)}. \tag{70}$$

The first observation is that exponential sums associated with $\mathbb{Z}_2$-valued quadratic forms can be computed recursively. Indeed, assume that $K_\alpha = 0$ for all α. Then

$$\mathcal{Z} = \sum_{x\in\{0,1\}^n} (-1)^{Q(x)} \quad \text{where} \quad Q(x) = xMx^T + Lx^T \pmod 2. \tag{71}$$

It will be convenient to consider more general quadratic forms $Q(x)$ as in Eq. (71) where M is an arbitrary binary matrix. We allow M to be non-symmetric and have non-zero diagonal.

Consider first the trivial case when M is a symmetric matrix. In this case all quadratic terms in $Q(x)$ cancel each other, that is, $Q(x)$ is linear. Thus $\mathcal{Z} = 2^n$ if $L = \mathrm{diag}(M)$ and $\mathcal{Z} = 0$ otherwise.

Suppose now that M is non-symmetric. We can assume wlog that $M_{1,2} \neq M_{2,1}$ (otherwise permute the variables). Then $M_{1,2} + M_{2,1} = 1 \pmod 2$. Write $x = (x_1, x_2, y)$ with $y \in \{0,1\}^{n-2}$. Define a partial sum

$$\mathcal{Z}(y) = \sum_{x_1,x_2 \in \{0,1\}} (-1)^{Q(x_1,x_2,y)} = \sum_{x_1,x_2 \in \{0,1\}} (-1)^{x_1 x_2 + \mu_1(y) x_1 + \mu_2(y) x_2 + Q_{else}(y)}, \tag{72}$$

where $Q_{else}(y)$ includes all terms in $Q(x)$ that do not depend on x_1, x_2,

$$\mu_1(y) = L_1 + M_{1,1} + \sum_{3 \leq \alpha \leq n} (M_{1,\alpha} + M_{\alpha,1}) y_\alpha \equiv L_1 + M_{1,1} + m_1 y^T,$$

$$\mu_2(y) = L_2 + M_{2,2} + \sum_{3 \leq \alpha \leq n} (M_{2,\alpha} + M_{\alpha,2}) y_\alpha \equiv L_2 + M_{2,2} + m_2 y^T.$$

Here m_1, m_2 are row vectors of length $n-2$. A simple algebra shows that

$$\sum_{x_1,x_2 \in \{0,1\}} (-1)^{x_1 x_2 + \mu_1 x_1 + \mu_2 x_2} = 2(-1)^{\mu_1 \mu_2} \qquad \text{for all } \mu_1, \mu_2 \in \{0,1\}. \tag{73}$$

Substituting this identity into Eq. (72) gives

$$\mathcal{Z} = \sum_{y \in \{0,1\}^{n-2}} \mathcal{Z}(y) = 2(-1)^{(L_1 + M_{1,1})(L_2 + M_{2,2})} \sum_{y \in \{0,1\}^{n-2}} (-1)^{Q'(y)}, \tag{74}$$

where $Q'(y)$ is a quadratic form that depends on $n-2$ variables:

$$Q'(y) = y(M_{else} + m_1^T m_2) y^T + (L_{else} + [L_1 + M_{1,1}] m_2 + [L_2 + M_{2,2}] m_1) y^T \tag{75}$$

The matrix M_{else} and the vector L_{else} are determined by $Q_{else}(y) = y M_{else} y^T + L_{else} y^T$. We have reduced the exponential sum problem with n variables to the one with $n-2$ variables. Clearly, the coefficients of $Q'(y)$ can be computed in time $O(n^2)$. The overall runtime is $\sum_{k=1}^{n} O(k^2) = O(n^3)$. This gives an algorithm for computing the exponential sum for a $\mathbb{Z}_2$-valued quadratic form.

Remark: The most time-consuming step is getting the matrix $M_{else} + m_1^T m_2$. Since the arithmetics is mod-2, this amounts to flipping all bits of M_{else} in a submatrix formed by rows $i \in m_1$ and by columns $j \in m_2$.

Consider now a $\mathbb{Z}_4$-valued form $q(x)$ defined in Eq. (69). Define a $\mathbb{Z}_2$-valued form

$$Q(x) = \sum_{1 \leq \alpha < \beta \leq n} (M_{\alpha,\beta} + K_\alpha K_\beta) x_\alpha x_\beta + \sum_{1 \leq \alpha \leq n} K_\alpha x_\alpha x_{n+1} + \sum_{1 \leq \alpha \leq n} L_\alpha x_\alpha \pmod 2. \tag{76}$$

Proposition 6. *Let $\mathcal{Z}$ be the exponential sum defined by Eqs. (69,70). Then*

$$\mathrm{Re}(\mathcal{Z}) = \frac{1}{2} \sum_{x \in \{0,1\}^{n+1}} (-1)^{Q(x)} \quad \textit{and} \quad \mathrm{Im}(\mathcal{Z}) = \frac{1}{2} \sum_{x \in \{0,1\}^{n+1}} (-1)^{Q(x) + x_{n+1}}. \tag{77}$$

Proof. Write $q(x) = 2r(x) + Kx^T \pmod 4$, where $r(x)$ is a $\mathbb{Z}_2$-valued quadratic form. Consider some $x \in \{0,1\}^n$ and let $\omega \equiv Kx^T \pmod 2$. One can easily check that

$$i^{Kx^T} = (-1)^{\sum_{1 \leq \alpha < \beta \leq n} K_\alpha K_\beta x_\alpha x_\beta} \cdot i^\omega.$$

By definition $\omega \in \{0,1\}$ so that

$$\mathrm{Re}(i^\omega) = \frac{1}{2}(1 + (-1)^\omega) \quad \text{and} \quad \mathrm{Im}(i^\omega) = \frac{1}{2}(1 - (-1)^\omega).$$

Define a $\mathbb{Z}_2$-valued form $Q'(x) = r(x) + \sum_{1\leq\alpha<\beta\leq n} K_\alpha K_\beta x_\alpha x_\beta$. Then

$$\mathrm{Re}(i^{q(x)}) = \frac{1}{2}\left[(-1)^{Q'(x)} + (-1)^{Q'(x)+Kx^T}\right] \quad \text{and} \quad \mathrm{Im}(i^{q(x)}) = \frac{1}{2}\left[(-1)^{Q'(x)} - (-1)^{Q'(x)+Kx^T}\right].$$

Finally, add an extra variable x_{n+1} such that the two terms in the square brackets correspond to $x_{n+1} = 0$ and $x_{n+1} = 1$ respectively. We arrive at Eq. (77) with $Q(x, x_{n+1}) = Q'(x) + x_{n+1}(Kx^T)$. □

□

Remark: Computing exponential sums associated with the real and imaginary parts of $\mathcal{Z}$ takes about the same time as computing a single exponential sum Eq. (71) because the forms $Q(x)$ and $Q(x) + x_{n+1}$ in Lemma 2 have the same quadratic parts.

Numerics shows that the new algorithm for computing exponential sums achieves a significant speedup as is shown in Table 1. Altogether, the use of the phase-sensitive Clifford simulator, sampling with equatorial states, and the improved Exponential Sum routine lead to a significant performance increase in simulations. In Table 2, we compare the performance of the simulator in Ref. [11] and this paper, when estimating the output probabilities of the Hidden Shift problem on 40-qubits with the Sum-over-Cliffords method (see also Sections 2.3 and 2.4).

Number of variables n	**10**	**20**	**30**	**40**	**50**	**60**
New runtime	0.016	0.017	0.021	0.023	0.030	0.036
BG16 runtime	0.42	0.50	0.77	1.10	1.40	1.72

Table 1: Average runtime in milliseconds of the new algorithm for computing exponential sums and comparison with the algorithm of Ref. [11]. Both simulations were performed on a Linux PC with a 3.2GHz Intel i5-6500 CPU.

Number of CCZ Gates	2	4	6
Number of states χ_Δ	39	149	497
New Runtime (s)	0.30	1.02	3.82
BG16 Runtime (s)	5.22	27.94	100.11

Table 2: Average runtime of the Norm Estimation step in seconds, for the new implementation compared with that of Ref. [11]. Norm Estimation is used to compute single qubit marginals on a 40-qubit state, with precision $\Delta = 0.3$. Both simulations were single-threaded, and run on a Linux PC with a 3.2GHz Intel i5-6500 CPU.

5 Stabilizer rank

In this Section, we describe bounds on the exact and approximate stabilizer rank. In subsection 5.1, we give the proof of Theorem 2, which proceeds by establishing an upper bound on the exact stabilizer rank of states symmetric under permutations of certain subsystems. As a consequence we will see that $\chi(\psi^{\otimes t}) \ll \chi(\psi)^t$ for modest t. In subsection 5.2 we prove Theorem 1 using a Sparsification lemma that allows us to convert exact stabilizer decompositions into approximate stabilizer decompositions (with possibly fewer terms). In subsection 5.3, we study the approximate stabilizer rank of Clifford magic states and establish Proposition 2. Finally, in subsection 5.4 we turn our attention to lower bounds and prove Proposition 3.

5.1 Exact stabilizer rank

Let us denote $\mathrm{Sym}_{n,t}$ as the subspace that is symmetric with respect to swaps between t partitions with each partition holding n qubits. For instance, any n-qubit state ψ satisfies $\psi^{\otimes t} \in \mathrm{Sym}_{n,t}$ for any t. Although the symmetric subspace also contains states entangled across these partitions. Throughout this section we use $\dim(\ldots)$ to denote the dimension of a vector space and $\mathrm{span}(\ldots)$ to denote the vector space spanned by a set of vectors. Let us agree that when we write $\dim(\mathbb{S})$

where $\mathbb{S}$ is a set of vectors (rather than a vector space) this means the dimension of the vector space spanned by $\mathbb{S}$.

This section provides a proof of Thm. 2, though we shall actually prove a more general result regarding the stabilizer rank of a subspace defined as follows

Definition 7. *We define stabilizer rank $\chi(P)$ of a subspace P to be the minimum χ such that there exists a set of χ stabilizer states $\mathbb{S} = \{\phi_1, \phi_2, \ldots, \phi_\chi\}$ satisfying $P \subset \mathrm{span}[\mathbb{S}]$.*

Notice that given a set of stabilizer states $\mathbb{S}$ such that $\mathrm{Sym}_{n,t} \subseteq \mathrm{span}(\mathbb{S})$, it follows that every element of the space $\mathrm{Sym}_{n,t}$ can be decomposed in terms of $|\mathbb{S}|$ stabilizer states. Therefore, if $\Psi \in \mathrm{Sym}_{n,t}$ then $\chi(\Psi) \leq \chi(\mathrm{Sym}_{n,t})$. As a special case, if $\Psi = \psi^{\otimes t}$ then $\chi(\psi^{\otimes t}) \leq \chi(\mathrm{Sym}_{n,t})$. Therefore, Thm. 2 follows as a corollary of the following result

Lemma 5. *Consider $\mathrm{Sym}_{n,t}$ for some nonzero n and t. It follows that for all $t \leq 5$ we have*

$$\chi(\mathrm{Sym}_{n,t}) = \dim[\mathrm{Sym}_{n,t}] = \binom{2^n + t - 1}{t}, \tag{78}$$

where the round brackets denotes the binomial coefficient.

This has the direct and elegant consequence that for all single qubit states ψ we have $\chi(\psi^{\otimes t}) \leq t + 1$ whenever $t \leq 5$.

Proof of Lemma 5. First we show that Eq. (78) holds for some n and t whenever there exists a set of stabilizer states $\mathbb{S}$ with the following properties:

1. every $\Phi \in \mathbb{S}$ satisfies $\Phi \in \mathrm{Sym}_{n,t}$; and
2. $\dim(\mathrm{Sym}_{n,t}) = \dim(\mathbb{S})$.

For any set of vectors $\mathbb{S}$, there exists a subset $\mathbb{S}' \subseteq \mathbb{S}$ that is a minimal spanning set, with $\mathrm{span}(\mathbb{S}') = \mathrm{span}(\mathbb{S})$ and $|\mathbb{S}'| = \dim(\mathbb{S})$. Therefore, given a set that spans the symmetric space we can conclude that $\chi(\mathrm{Sym}_{n,t}) \leq \dim(\mathbb{S})$. Furthermore, if $\mathbb{S}$ has the swap invariance property then $\mathrm{span}(\mathbb{S}) \subseteq \mathrm{Sym}_{n,t}$ and $\dim(\mathbb{S}) \leq \dim(\mathrm{Sym}_{n,t})$. Combining these inequalities gives $\chi(\mathrm{Sym}_{n,t}) \leq \dim(\mathrm{Sym}_{n,t})$. It is obvious that $\dim(\mathrm{Sym}_{n,t}) \leq \chi(\mathrm{Sym}_{n,t})$ and so $\chi(\mathrm{Sym}_{n,t}) = \dim(\mathrm{Sym}_{n,t})$. Lastly, the dimension of the symmetric space is well-known and can for example be found in Ref. [59].

Next, it remains to find a set $\mathbb{S}$ with the aforementioned properties for certain values of n and t. We consider sets of stabilizer states of the form $\mathbb{S}_{n,t} = \{|\phi_j\rangle^{\otimes t}\}_j$ where $\{|\phi_j\rangle\}_j =: \mathrm{STAB}_n$ is the set of all n-qubit stabilizer states. This ensures property 1. It remains to show when $\mathbb{S}_{n,t}$ has sufficiently large dimension (property 2). We observe that the operator

$$\sigma_{n,t} := \frac{1}{|\mathrm{STAB}_n|} \sum_{\psi_j \in \mathrm{STAB}_n} |\psi_j\rangle\langle\psi_j|^{\otimes t} \tag{79}$$

satisfies

$$\mathrm{rank}(\sigma_{n,t}) = \dim(\mathbb{S}_{n,t}). \tag{80}$$

and so property 2 also holds whenever $\mathrm{rank}(\sigma_{n,t}) = \dim(\mathrm{Sym}_{n,t})$.

Let us consider when $t \leq 3$ with no constraints on n. We will use that the stabilizer states form a projective 3-design [38, 57, 59]. The relevant property of such designs is that for $t \leq 3$ we know

$$\sigma_{n,t} \propto \Pi_{n,t}, \tag{81}$$

where $\Pi_{n,t}$ is the projector onto $\mathrm{Sym}_{n,t}$. Therefore, $\mathrm{rank}(\sigma_{n,t}) = \mathrm{rank}(\Pi_{n,t}) = \dim(\mathrm{Sym}_{n,t})$ and the lemma is proven for the case of $t \leq 3$.

For $t = 4$, it is known that the stabilizer states are not a projective 4-design and so $\sigma_{n,4}$ is not proportional to the symmetric projector [59]. However, the stabilizer states "fail gracefully" to be a projective 4-design [59], such that the deviation of $\sigma_{n,4}$ from $\Pi_{n,4}$ is sufficiently small that we still have $\mathrm{rank}(\sigma_{n,4}) = \mathrm{rank}(\Pi_{n,4})$. Ref. [29] extends this result such that we can also deduce the following

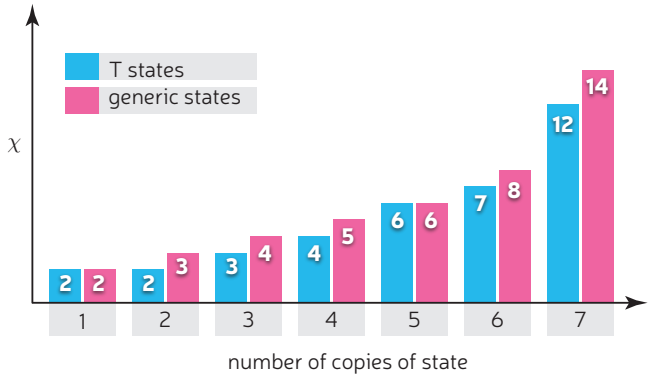


Figure 6: The exact stabilizer rank (numerically found) for n copies of a single qubit state: for the T state and for generic single qubit states.

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
$n=1$	2	1.73205	1.5874	1.49535	1.43097
$n=2$	4	3.16228	2.71442	2.4323	2.23685
$n=3$	8	6	4.93242	4.26215	3.79966

Table 3: Upper bounds on $\chi(\psi^{\otimes t})^{1/t}$ where ψ is an n qubit state. Asymptotically we have $\chi(\psi^{\otimes N}) \leq (\chi(\psi^{\otimes t})^{1/t})^N$. Since lower values lead to lower simulation overhead we see a significant advantage in using blocks of size up to 5.

Claim 1. *For all n and $t \leq 5$ we have* $\mathrm{rank}(\sigma_{n,t}) = \mathrm{rank}(\Pi_{n,t})$.

This suffices to prove Lem. 5. In contrast, this proof technique can not extend to $t > 5$ due to the stabilizer testing algorithm of Ref [29]. This algorithm shows that there exists a projector W such that $\mathrm{Tr}[W\sigma_{n,6}] = 0$ but $\mathrm{Tr}[W\Pi_{n,6}] \neq 0$, which entails $\mathrm{rank}(\sigma_{n,6}) < \mathrm{rank}(\Pi_{n,6})$.

Although Claim 1 can be deduced from Ref. [29], it is not explicitly shown, so we provide the details here. Examples 4.27 and 4.28 of Ref. [29], show that

$$\begin{aligned} \sigma_{n,4} &\propto \Pi_{n,4} + a_n \Pi_{n,4} P_{[4]}^{\otimes n} \Pi_{n,4}, \\ \sigma_{n,5} &\propto \Pi_{n,5} + b_n \Pi_{n,5} P_{[5]}^{\otimes n} \Pi_{n,5}. \end{aligned} \tag{82}$$

where a_n and b_n are positive constants and $P_{[4]}$ and $P_{[5]}$ are projectors onto a stabilizer code

$$\begin{aligned} P_{[4]} &= \frac{1}{4}(\mathbb{1}^{\otimes 4} + X^{\otimes 4} + Y^{\otimes 4} + Z^{\otimes 4}) \\ P_{[5]} &= P_{[4]} \otimes \mathbb{1} \end{aligned} \tag{83}$$

Since $P_{[4]}$ and $P_{[5]}$ are positive operators, so too are $a_n \Pi_{n,4} P_{[4]}^{\otimes n} \Pi_{n,4}$ and $b_n \Pi_{n,5} P_{[5]}^{\otimes n} \Pi_{n,5}$. In general, if M and N are positive operators we have $\mathrm{rank}(M+N) \geq \mathrm{rank}(M)$. Therefore, for $t = 4, 5$ we have $\mathrm{rank}(\sigma_{n,t}) \geq \mathrm{rank}(\Pi_{n,t})$, which implies the desired rank equivalence and completes the proof. □

We reflect that we have proved Lem. 5, from which Thm. 2 follows immediately. For single qubit states ($n = 1$) this entails that

$$\chi(\psi^{\otimes t}) \leq t + 1, \forall t \leq 5. \tag{84}$$

The rest of this subsection discusses numerical experiments into whether this inequality is tight.

Clearly the bound is loose for stabilizer states since then we have $\chi(\psi^{\otimes t}) = 1 < t+1$. However, Clifford magic states are also exceptional for many t values. Bravyi, Smith and Smolin [14] discuss the stabilizer rank of single qubit states that are an eigenstate of some Clifford unitary. For instance, the $|T\rangle$ Clifford magic states are exceptional in that for $2 \leq t \leq 4$ we have that $\chi(T^{\otimes t}) = t < t+1$,

which we illustrate in Fig. 6. We remark that $|T\rangle$ has the Clifford symmetry $C_T|T\rangle = |T\rangle$ for $C_T = TXT^\dagger$. In total there are 12 single qubit states in the Clifford orbit of $|T\rangle$. An additional class of Clifford symmetric states is the Clifford orbit of the face state $|f\rangle$

$$|f\rangle\langle f| = \frac{1}{2}\left(\mathbb{1} + \frac{X+Y+Z}{\sqrt{3}}\right), \tag{85}$$

which comprises 8 different states. The face state is an eigenstate of the Clifford $C_F = e^{-i\pi/12}SH$ that cyclically permutes Pauli X,Y and Z. Bravyi, Smith and Smolin reported (see conjecture 1 of Ref. [14]) that $\chi(f^{\otimes t})$ appears to equal $\chi(T^{\otimes t})$, providing another class of states where Eq. (84) is not tight.

Next, we ask if there are any other single qubit states for which Eq. (84) is not tight. We proceed by a heuristic, numerical search, extending the search method of Ref. [14]. To find a decomposition of a state $|\psi\rangle$, we use an objective function $F_\Psi\left(\{|\phi_j\rangle\}\right) = ||\Pi|\Psi\rangle||$ where Π is a projector onto span $\left(\{|\phi_j\rangle\}\right)$. We start by choosing a set of k random stabilizer states $\{|\phi_j\rangle\}$, with $k = 2$ on the first run. Random stabilizer states were obtained by generating a random binary matrix, using the algorithm of Garcia et al. to convert it to a canonical stabilizer tableau, and computing the corresponding state vector [25]. Let the value of the objective function at a given timestep be F. We update one stabilizer state in the set by applying a random Pauli projector, and evaluate the objective function on the new set $F_\Psi\left(\{|\phi_j\rangle\}'\right) = F'$. If $F' > F$ then we accept the move, otherwise the new decomposition is accepted with a probability $p = \exp\left[-\beta\left(F - F'\right)\right]$, where β is an inverse temperature parameter that decreases as the walk proceeds [14]. If F equals 1 at any point in the walk, we halt and conclude $\chi(\Psi) \leq k$. If F does not converge to unity within a constant number of steps, we increment k and start again.

Random typical states were generated as $|\psi\rangle = U|0\rangle$, where U are Haar random unitaries. We sampled 1000 Harr random states and numerically estimated the stabilizer rank of $\Psi = \psi^{\otimes t}$ using the above method. In every instance, the best decomposition we found saturated the inequalities of Eq. (84). We also examined conjecture 1 of [14], by searching for decompositions of single-qubit Clifford magic states. All decompositions found were below the bound of Eq. (84).

Although these numerical searches were not exhaustive, the results support the hypothesis that Eq. (84) is an equality for typical single qubit states. This supports the conjecture that Eq. (84) is tight, if and only if the state has no Clifford symmetries.

As a closing remark, we comment on consequences of these results for simulation overheads. If a circuit contains many copies of the same multi-qubit phase gate, simulation overheads are reduced by working with blocks of magic states as shown in Table. 3.

5.2 Sparsification Lemma

Our new bounds on the approximate stabilizer rank in Theorem 1 are obtained using the following lemma. It shows how to convert a stabilizer decomposition of some target state ψ with a small l_1 norm to a sparse stabilizer decomposition of ψ.

Lemma 6 (Sparsification). *Let ψ be a normalized n-qubit state with a decomposition $|\psi\rangle = \sum_j c_j|\phi_j\rangle$ where all ϕ_j are normalized stabilizer states and $c_j \in \mathbb{C}$. For any integer k there exists a distribution of random quantum states $|\Omega\rangle$ of the form $|\Omega\rangle = \frac{\|c\|_1}{k}\sum_{\alpha=1}^{k}|\omega_\alpha\rangle$ where each $|\omega_\alpha\rangle$ is (up to a global phase) one of the states $\{|\phi_j\rangle\}$ and*

$$\mathbb{E}\left(\|\psi - \Omega\|^2\right) = \frac{\|c\|_1^2}{k}, \tag{86}$$

where $\|c\|_1 := \sum_j |c_j|$ and $\|\psi\| = \sqrt{\langle\psi|\psi\rangle}$.

Theorem 1 is a simple corollary of Lemma 6. Indeed, assume that all ϕ_j are stabilizer states. Choosing $k = (\|c\|_1/\delta)^2$ we find that the right-hand side is upper-bounded by δ^2. Therefore there exists at least one $|\Omega\rangle$ (which is manifestly a sum of k stabilizer states) that δ-approximates $|\psi\rangle$. This proves Theorem 1.

Note that we can use Markov's inequality and Eq. (86) to lower bound the probability that a randomly chosen Ω is a good approximation to ψ, e.g.,

$$\Pr\left[\|\psi - \Omega\|^2 \geq 2\delta^2\right] \leq 1/2 \quad \text{for} \quad k \geq \frac{\|c\|_1^2}{\delta^2}.$$

Suppose that we randomly choose some $|\Omega\rangle$ as prescribed above. Can we estimate how well it approximates ψ? The following Lemma can be used for this purpose.

Lemma 7 (**Sparsification tail bound**). *Let ψ, Ω, k be as in Lemma 6. If we choose $k \geq \frac{\|c\|_1^2}{\delta^2}$ then*

$$\mathbb{E}\left[\langle\Omega|\Omega\rangle - 1\right] \leq \delta^2, \tag{87}$$

and

$$\Pr\left[\|\psi - \Omega\|^2 \leq \langle\Omega|\Omega\rangle - 1 + \delta^2\right] \geq 1 - 2\exp\left(-\frac{\delta^2}{8F(\psi)}\right). \tag{88}$$

Note that we are interested in cases where the stabilizer fidelity $F(\psi)$ is exponentially small as a function of the number of qubits n. In such cases the Lemma states that

$$\|\psi - \Omega\|^2 \leq \langle\Omega|\Omega\rangle - 1 + \delta^2,$$

with all but vanishingly small probability if n is sufficiently large. Moreover, the quantity $\langle\Omega|\Omega\rangle$ appearing in the above can be approximated to a given relative error using the norm estimation algorithm from Section 4.3 which has runtime scaling linearly with k.

Proof of Lemma 6. Define a probability distribution $p_j := |c_j|/||c||_1$ and write

$$|\psi\rangle = \|c\|_1 \sum_j p_j |W_j\rangle \tag{89}$$

where $|W_j\rangle := (c_j/|c_j|)|\phi_j\rangle$ are normalized stabilizer states. Now define a random variable $|\omega\rangle$ which is equal to $|W_j\rangle$ with probability p_j. Then

$$|\psi\rangle = \|c\|_1 \mathbb{E}\left[|\omega\rangle\right]. \tag{90}$$

Let k be a positive integer and consider a random state

$$|\Omega\rangle = \frac{\|c\|_1}{k} \sum_{\alpha=1}^{k} |\omega_\alpha\rangle, \tag{91}$$

where $\omega_1, \omega_2, \ldots, \omega_k$ are i.i.d random copies of $|\omega\rangle$. By construction, on average we have

$$\mathbb{E}[\langle\psi|\Omega\rangle] = \mathbb{E}[\langle\Omega|\psi\rangle] = 1 \tag{92}$$

even though for any particular random sample $\langle\Omega|\psi\rangle \neq 1$. In general, not only will Ω not be proportional to ψ, but Ω will not be correctly normalized. However, the normalization can be bounded in expectation as follows

$$\mathbb{E}\left[\langle\Omega|\Omega\rangle\right] = \frac{\|c\|_1^2}{k^2}\mathbb{E}\left[\sum_{\alpha=1}^{k}\langle\omega_\alpha|\omega_\alpha\rangle\right] + \frac{\|c\|_1^2}{k^2}\mathbb{E}\left[\sum_{\alpha\neq\beta}\langle\omega_\alpha|\omega_\beta\rangle\right] \tag{93}$$

$$= \|c\|_1^2 \frac{\mathbb{E}\left[\langle\omega|\omega\rangle\right]}{k} + \frac{1}{k^2}k(k-1) \tag{94}$$

$$\leq 1 + \frac{\|c\|_1^2}{k} \tag{95}$$

where in the second line we used the fact that $\|c\|_1^2\mathbb{E}\left[\langle\omega_\alpha|\omega_\beta\rangle\right] = \langle\psi|\psi\rangle$ for $\alpha \neq \beta$.

We are interested in the expected error

$$\mathbb{E}\left[\||\psi\rangle - |\Omega\rangle\|^2\right] = \mathbb{E}\left[\langle\Omega|\Omega\rangle\right] - \mathbb{E}\left[\langle\Omega|\psi\rangle\right] - \mathbb{E}\left[\langle\psi|\Omega\rangle\right] + \mathbb{E}\left[\langle\psi|\psi\rangle\right] \tag{96}$$

Using $\langle\psi|\psi\rangle = 1$, Eq. (95) and Eq. (92) we find

$$\mathbb{E}\left[\||\psi\rangle - |\Omega\rangle\|^2\right] = \frac{\|c\|_1^2}{k}. \tag{97}$$

This completes the proof of Lemma 6. □

Proof of Lemma 7. Equation (87) follows directly from Eq. (95) and the choice of k. Define random variables

$$X_\alpha = \|c\|_1 \mathrm{Re}(\langle\psi|\omega_\alpha\rangle) \qquad 1 \le \alpha \le k$$

and let

$$\bar{X} = \frac{1}{k}\sum_{\alpha=1}^{k} X_\alpha = \mathrm{Re}(\langle\psi|\Omega\rangle).$$

Then

$$|\mathrm{Re}(\langle\psi|\Omega\rangle) - 1| = \left|\bar{X} - \mathbb{E}[\bar{X}]\right|. \tag{98}$$

Now $\bar{X}$ is a sample mean of k independent and identically distributed random variables X_α, each of which is bounded as

$$|X_\alpha| \le \|c\|_1 |\langle\psi|\omega_\alpha\rangle| \le \|c\|_1 \sqrt{F(\psi)} \tag{99}$$

where in the last inequality we used the definition of stabilizer fidelity. Applying Hoeffding's inequality [31] and using Eqs. (98, 99) gives

$$\Pr\left[|\mathrm{Re}(\langle\psi|\Omega\rangle) - 1| \ge \frac{\delta^2}{2}\right] \le 2\exp\left(-\frac{2k\delta^4}{4\left(2\|c\|_1\sqrt{F(\psi)}\right)^2}\right) \le 2\exp\left(-\frac{\delta^2}{8F(\psi)}\right) \tag{100}$$

where we used $k \ge \|c\|_1^2/\delta^2$. Finally, applying the triangle inequality to Eq. (96) gives

$$\|\psi - \Omega\|^2 \le \langle\Omega|\Omega\rangle - 1 + 2\,|1 - \mathrm{Re}(\langle\psi|\Omega\rangle)| \tag{101}$$

Combining Eqs. (101, 100) completes the proof.

□

5.3 Approximate stabilizer rank of Clifford magic states

Proposition 2 asserts that $\xi(\psi) = F(\psi)^{-1}$ when ψ is a Clifford magic state (Def 5). In fact, this relation holds for a wider class of ψ and we comment on this at the end of the following proof. Recall that a Clifford magic state ψ is stabilized by a group of Clifford unitaries with generators $Q_j := VX_jV^\dagger$. We denote this group as $\mathcal{Q} := \langle Q_j\rangle = \langle VX_jV^\dagger\rangle$. Here we describe upper bounds on the approximate stabilizer rank of Clifford magic states. We begin with the proof of Proposition 2

Proof of Proposition 2. From the definition of Clifford magic states, we have

$$\begin{aligned} P_\psi = |\psi\rangle\langle\psi| &= V\frac{1}{2^n}\prod_j(\mathbb{1} + X_j)V^\dagger \\ &= \frac{1}{2^n}\prod_j(\mathbb{1} + Q_j) \\ &= \frac{1}{|\mathcal{Q}|}\sum_{q\in\mathcal{Q}} q \end{aligned} \tag{102}$$

Let ϕ_0 be a stabilizer state such that $|\langle\psi|\phi_0\rangle|^2 > 0$. Then

$$\begin{aligned} |\psi\rangle &= \frac{|\psi\rangle\langle\psi|\phi_0\rangle}{\langle\psi|\phi_0\rangle} \\ &= \left[\frac{1}{|\mathcal{Q}|}\sum_{q\in\mathcal{Q}} q\right]\frac{|\phi_0\rangle}{\langle\psi|\phi_0\rangle} \\ &= \frac{1}{|\mathcal{Q}|\langle\psi|\phi_0\rangle}\sum_{q\in\mathcal{Q}} q|\phi_0\rangle. \end{aligned} \tag{103}$$

Using Eq. (103) and the fact that $q|\phi_0\rangle$ is a stabilizer state for all $q \in \mathcal{Q}$ we immediately obtain

$$||c||_1^2 = \frac{1}{|\langle\psi|\phi_0\rangle|^2},$$

for this decomposition. To minimise $||c||_1^2$ it is natural to use the stabilizer state with the larger possible overlap, $F(\psi) = \max_{\phi_0} |\langle\psi|\phi_0\rangle|^2$, which we call the stabilizer fidelity. Therefore, once we have found a ϕ_0 attaining the maximum, we have a decomposition achieving $||c||_1^2 = F(\psi)^{-1}$. This discussion suffices to prove that

$$\xi(\psi) \leq F(\psi)^{-1}.$$

To establish the converse consider any stabilizer decomposition

$$|\psi\rangle = \sum_{j=1}^{\chi} c_j |\phi_j\rangle.$$

Taking the inner product with ψ we get

$$1 = \left| \sum_{j=1}^{\chi} c_j \langle\psi|\phi_j\rangle \right| \leq \|c\|_1 \sqrt{F(\psi)},$$

where we used the fact that $|\langle\psi|\phi_j\rangle|^2 \leq F(\psi)$. Squaring the above completes the proof.

More generally, let $\mathcal{Q}$ be *any* subgroup of the Clifford group satisfying $|\psi\rangle\langle\psi| = |\mathcal{Q}|^{-1} \sum_{q\in\mathcal{Q}} q$ and with exactly one group element (the identity) stabilizing $|\phi_0\rangle$. The above proof goes through unmodified, but admits a wider class of states for which $\xi(\psi) = F(\psi)^{-1}$ including the face state, $|f\rangle$, satisfying

$$|f\rangle\langle f| = \frac{1}{2}\left(\mathbb{1} + \frac{(X+Y+Z)}{\sqrt{3}}\right) = \frac{1}{|\mathcal{Q}|}\sum_{q\in\mathcal{Q}} q \tag{104}$$

where $\mathcal{Q} = \{\mathbb{1}, C_F, C_F^2\}$ and $C_F = e^{-i\pi/12} SH$ is the Clifford that cyclically permutes Pauli X,Y and Z. □

The $|T\rangle^{\otimes n}$ state is the most well known example of a Clifford magic state. It has been shown (see Lemma 2 of Ref. [16] or Lemma 2 of Ref. [11]) that $F(T^{\otimes n})^{-1} = |\langle +|T\rangle|^{2n}$ and so $|+\rangle^{\otimes n}$ can be used to generate the decomposition with optimal $\xi(\psi)$. Combining this with Lemma 6 gives the same upper bound on $\chi_\delta(T^{\otimes n})$ as was previously shown in Ref. [11]. However, the techniques are slightly different. Our Lemma 6 randomly selects a subset of terms from the decomposition, whereas Ref. [11] randomly select a subset of terms that form a random linear code. We remark that the random linear code construction also generalises to all Clifford magic states. For any linear code $\mathcal{L} \subseteq \mathbb{F}_2^n$ we can associate a subgroup $\mathcal{Q}_\mathcal{L} \subseteq \mathcal{Q}$. That is, given a decomposition as in Eq. (103) with group $\mathcal{Q}$, we can choose a random subgroup $\mathcal{Q}_\mathcal{L} \subseteq \mathcal{Q}$ and define the normalised approximate state

$$|\mathcal{L}\rangle \propto \sum_{q\in\mathcal{Q}_\mathcal{L}} q|\phi_0\rangle. \tag{105}$$

Following analogous steps to those in Ref. [11], one can show that this approach gives the same asymptotic scaling of χ_δ as in Lemma 6. While the behaviour of χ_δ is identical, it may be easier to implement a simulator working with random subgroups than random subsets.

As a further example, let us consider the Clifford magic state corresponding to a CCZ (control-control-Z) gate,

$$|CCZ\rangle = CCZ|+\rangle|+\rangle|+\rangle = \frac{1}{\sqrt{8}} \sum_{a,b,c\in\{0,1\}} (-1)^{abc}|a\rangle|b\rangle|c\rangle \tag{106}$$

This magic state is the "+1" eigenstate for a group $\mathcal{Q}$ with three generators of the form $CCZ \cdot X_j \cdot CCZ^\dagger$. More explicitly these generators are

$$\begin{aligned} Q_1 &= CCZ \cdot X_1 \cdot CCZ^\dagger = X_1 CZ_{2,3} \\ Q_2 &= CCZ \cdot X_2 \cdot CCZ^\dagger = X_2 CZ_{1,3} \\ Q_3 &= CCZ \cdot X_3 \cdot CCZ^\dagger = X_3 CZ_{1,2} \end{aligned} \tag{107}$$

where $CZ_{i,j}$ denotes a control-Z between qubits i and j. One can straightforwardly confirm that $F(CCZ) = |\langle +++|CCZ\rangle|^2 = 9/16$, and that

$$|CCZ\rangle = \frac{1}{6}\sum_{Q\in\mathcal{Q}} Q|++\rangle, \tag{108}$$

has $||c||_1^2 = 16/9$. Using this decomposition for many CCZ states shows $\chi_\delta(CCZ^{\otimes t}) \leq \delta^{-2}(9/16)^t \sim \delta^{-2}1.778^t$. Note that this is slower exponential scaling than obtained by synthesizing each CCZ with 4 T-gates and using $\chi_\delta(T^{\otimes 4t}) \leq \delta^{-2}1.884^t$. It is conceivable that a better decomposition exists since ξ only provides an upper bound on the approximate stabilizer rank.

One could obtain better decompositions if the stabilizer fidelity is not multiplicative, but we show later (see Corollary 3) that $F(T^{\otimes t}) = F(T)^t$ and $F(CCZ^{\otimes t}) = F(CCZ)^t$. However, one of the significant open questions remaining from this work is whether stabilizer fidelity is always multiplicative for all Clifford magic states. Lastly, we remark that one can lift the above stabilizer decomposition to obtain a Clifford unitary decomposition of CCZ that can be used for an approximate sum-over-Cliffords simulator.

5.4 Lower bound based on ultra-metric matrices

Previous sections give explicit stabilizer decompositions of states and therefore upper bounds on the stabilizer rank. Yet we have no techniques that provide lower bounds on the stabilizer rank that scale exponentially with the number of copies. Here we present results in this direction. Let $|H\rangle = \cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle$ be the magic state which is Clifford equivalent to $|T\rangle$. We would like to approximate n copies of $|H\rangle$ by a low-rank linear combination of stabilizer states

$$|\tilde{x}\rangle = |\tilde{x}_1\rangle \otimes \cdots \otimes |\tilde{x}_n\rangle \quad \text{where} \quad |\tilde{0}\rangle = |0\rangle \quad \text{and} \quad |\tilde{1}\rangle = |+\rangle.$$

Here we derive a lower bound on the rank of such approximations stated earlier as Prop. 3. We first restate this result as follows

Theorem 3. *Suppose $S \subseteq \{0,1\}^n$ is an arbitrary subset and ϕ is an arbitrary linear combination of states $|\tilde{x}\rangle$ with $x \in S$ such that $\|\phi\| = 1$. Then*

$$|S| \geq |\langle H^{\otimes n}|\phi\rangle|^2 \cdot \cos(\pi/8)^{-2n}. \tag{109}$$

Proof. Let $\chi = |S|$ and $S = \{x^1, x^2, \ldots, x^\chi\}$ for some bit strings x^i. The orthogonal projector onto a linear subspace spanned by the states $|\tilde{x}^1\rangle, \ldots, |\tilde{x}^\chi\rangle$ has the form

$$\Pi = \sum_{i,j=1}^{\chi} (G^{-1})_{i,j} |\tilde{x}^i\rangle\langle\tilde{x}^j|, \tag{110}$$

where G is the Gram matrix defined by $G_{i,j} = \langle\tilde{x}^i|\tilde{x}^j\rangle = t^{|x^i \oplus x^j|}$, with $t = 2^{-1/2}$. Here and below $\oplus$ denotes addition of bit strings modulo two. Noting that $\langle\tilde{x}|H^{\otimes n}\rangle = \cos(\pi/8)^n$ for all x one gets

$$|\langle H^{\otimes n}|\phi\rangle|^2 \leq \langle H^{\otimes n}|\Pi|H^{\otimes n}\rangle = \cos(\pi/8)^{2n} \sum_{i,j=1}^{\chi} (G^{-1})_{i,j} \leq \chi \cos(\pi/8)^{2n}. \tag{111}$$

The last inequality follows from

Lemma 8. *Suppose $x^1, \ldots, x^\chi \in \{0,1\}^n$ are distinct bit strings and $0 < t < 1$ is a real number. Let G be a matrix of size χ with entries*

$$G_{i,j} = t^{|x^i \oplus x^j|}. \tag{112}$$

Then G is invertible and

$$\sum_{i,j=1}^{\chi} (G^{-1})_{i,j} \leq \chi. \tag{113}$$

Proof. Let $|1\rangle, |2\rangle, \ldots, |\chi\rangle$ be the basis vectors of $\mathbb{R}^\chi$ such that $G_{i,j} = \langle i|G|j\rangle$. We claim that Eq. (113) holds whenever one can find a family of matrices G_σ and probabilities $p_\sigma \geq 0$ such that

(a) $G = \sum_\sigma p_\sigma G_\sigma$ and $\sum_\sigma p_\sigma = 1$

(b) G_σ is positive definite

(c) $0 \leq \langle i|G_\sigma|j\rangle \leq 1$ and $\langle i|G_\sigma|i\rangle = 1$

(d) $\langle i|G_\sigma^{-1}|j\rangle \leq 0$ for $i \neq j$

Indeed, let $|e\rangle$ be the all-ones vector, $|e\rangle = \sum_{i=1}^\chi |i\rangle$. We have to prove that $\langle e|G^{-1}|e\rangle \leq \chi$. Conditions (a,b) imply that G is positive definite (and thus invertible). Noting that the function $f(x) = x^{-1}$ is operator convex on the interval $(0, \infty)$ one gets

$$\langle e|G^{-1}|e\rangle \leq \sum_\sigma p_\sigma \langle e|G_\sigma^{-1}|e\rangle. \tag{114}$$

From conditions (c,d) one gets

$$\langle i|G_\sigma^{-1}|j\rangle \leq \langle i|G_\sigma^{-1}|j\rangle\langle j|G_\sigma|i\rangle$$

for $i \neq j$ with the equality for $i = j$. Therefore

$$\langle e|G_\sigma^{-1}|e\rangle = \sum_{i,j=1}^{\chi} \langle i|G_\sigma^{-1}|j\rangle \leq \sum_{i,j=1}^{\chi} \langle i|G_\sigma^{-1}|j\rangle\langle j|G_\sigma|i\rangle = \mathrm{Tr}(G_\sigma^{-1}G_\sigma) = \mathrm{Tr}(I) = \chi. \tag{115}$$

Substituting this into Eq. (114) gives $\langle e|G^{-1}|e\rangle \leq \chi \sum_\sigma p_\sigma = \chi$, as desired.

It remains to construct the requisite matrices G_σ. Our construction is based on the so-called *ultrametric matrices*, see Refs. [41, 45].

Definition 8. *A symmetric real matrix A is called ultrametric iff $0 \leq A_{i,j} < 1$ for $i \neq j$, $A_{i,i} = 1$, and*

$$A_{i,j} \geq \min\left(A_{i,k}, A_{j,k}\right) \quad \textit{for all } i,j,k. \tag{116}$$

The last condition demands that for any triple of elements $A_{i,j}$, $A_{i,k}$, $A_{j,k}$ the two smallest elements coincide. The following fact was established in Refs. [41, 45].

Fact 1. *Suppose A is an ultrametric matrix. Then A is invertible and positive definite. Furthermore, $\langle i|A^{-1}|j\rangle \leq 0$ for all $i \neq j$.*

Thus it suffices to show that G is a probabilistic mixture of ultrametric matrices. Indeed, if condition (a) holds for some ultrametric matrices G_σ then condition (c) follows directly from Definition 8 while conditions (b,d) follow from Fact 1.

The first step is to equip the Boolean cube $\{0,1\}^n$ with a distance function that obeys an analogue of the ultrametricity condition Eq. (116). Given a pair of bit strings $x, y \in \{0,1\}^n$, define $d(x,y)$ as the smallest integer $j \geq 0$ such that the last $n-j$ bits of x and y coincide (that is, $x_i = y_i$ for all $i > j$). We set $d(x,y) = n$ if $x_n \neq y_n$. Note that $d(x,y)$ is different from the Hamming distance. For example, $d(101, 111) = 2$ and $d(101, 100) = 3$. By definition $d(x,y) \in [0,n]$ and $d(x,y) = 0$ iff $x = y$. Furthermore, $d(x,y)$ depends only on $x \oplus y$. We claim that

$$d(x,y) \leq \max\left\{d(x,z), d(z,y)\right\} \tag{117}$$

for any triple of strings x, y, z. Indeed, let $j = \max\{d(x,z), d(z,y)\}$. Then $x_i = z_i = y_i$ for all $i > j$, that is, $d(x,y) \le j$.

Suppose q_w is a normalized probability distribution on the set of integers $w = 0, 1, \ldots, n$ such that $q_w > 0$ for all w. Define a $\chi \times \chi$ matrix A such that

$$A_{i,j} = \sum_{w \ge d(x^i, x^j)} q_w. \tag{118}$$

Here x^i and x^j are the bit strings from the statement of the lemma. We claim that A is ultrametric (according to Definition 8). Indeed, consider any triple i, j, k as in Eq. (116) and assume wlog that $A_{i,k} \le A_{j,k}$. Since the matrix element $A_{i,j}$ is a monotone decreasing function of the distance $d(x^i, x^j)$, we get $d(x^i, x^k) \ge d(x^j, x^k)$. Then Eq. (117) gives $d(x^i, x^j) \le d(x^i, x^k)$. Using the monotonicity again one gets $A_{i,j} \ge A_{i,k} = \min\{A_{i,k}, A_{j,k}\}$, confirming Eq. (116). The remaining conditions $0 \le A_{i,j} < 1$ for $i \ne j$ and $A_{i,i} = 1$ follow from the assumption that all bit strings x^i are distinct and that q_w is a normalized probability distribution. Thus the matrix A defined by Eq. (118) is indeed ultrametric.

We are now ready to define a family of ultrametric matrices G_σ such that $G = \sum_\sigma p_\sigma G_\sigma$. Let us choose the label σ as a permutation of n integers, $\sigma \in S_n$. The distribution p_σ will be the uniform distribution on the symmetric group, that is, $p_\sigma = 1/n!$ for all $\sigma \in S_n$. Given a permutation σ and a bit string $x \in \{0,1\}^n$ let $\sigma(x) \in \{0,1\}^n$ be the result of permuting bits of x according to σ. We set

$$\langle i|G_\sigma|j\rangle = \sum_{w \ge d(\sigma(x^i), \sigma(x^j))} q_w. \tag{119}$$

The same argument as above confirms that G_σ is ultrametric for any permutation σ. Define

$$G' = \frac{1}{n!} \sum_{\sigma \in S_n} G_\sigma. \tag{120}$$

We claim that $\langle i|G'|j\rangle = \langle i|G|j\rangle = t^{|x^i \oplus x^j|}$ for a suitable choice of probabilities q_w. Indeed, the identity $d(x,y) = d(0^n, x \oplus y)$ implies that a matrix element $\langle i|G_\sigma|j\rangle$ depends only on $x^i \oplus x^j$. By the symmetry, matrix elements $\langle i|G'|j\rangle$ depend only on the Hamming weight $h = |x^i \oplus x^j|$. Therefore it suffices to compute $\langle i|G'|j\rangle$ for the special case when $x^i = 0^n$ is the all-zero string and x^j is any fixed bit string with the Hamming weight h, for example, $x^j = 1^h 0^{n-h}$. Then

$$\langle i|G'|j\rangle = \frac{1}{n!} \sum_{\sigma \in S_n} \sum_{w \ge d(0^n, \sigma(1^h 0^{n-h}))} q_w. \tag{121}$$

By definition of the distance $d(x,y)$ one gets $d(0^n, \sigma(1^h 0^{n-h})) \le w$ iff $h \le w$ and $\sigma_1, \ldots, \sigma_h \le w$. The number of such permutations σ is $\binom{w}{h} h!(n-h)!$. Exchanging the sums over σ and w in Eq. (121) one gets

$$\langle i|G'|j\rangle = \frac{1}{n!} \sum_{w=h}^{n} \binom{w}{h} h!(n-h)!\, q_w. \tag{122}$$

We shall choose q_w as a binomial distribution,

$$q_w = \binom{n}{w} t^w (1-t)^{n-w}. \tag{123}$$

Substituting Eq. (123) into Eq. (122) and introducing a variable $p = w - h$ one gets

$$\langle i|G'|j\rangle = \sum_{p=0}^{n-h} \binom{n-h}{p} t^{p+h} (1-t)^{n-h-p} = t^h. \tag{124}$$

By definition, $h = |x^i \oplus x^j|$, so that $G' = G$ as claimed. Thus G is indeed a probabilisitic mixture of ultrametric matrices and the lemma is proved. □

□

6 Stabilizer fidelity and Stabilizer extent

In the previous Section we established upper bounds on the approximate stabilizer rank of a state ψ which depend on the the squared 1-norm $\|c\|_1^2$, where

$$|\psi\rangle = \sum_j c_j |\phi_j\rangle,$$

is a given stabilizer decomposition. Recall that the stabilizer extent $\xi(\psi)$ denotes the minimum value of $||c||_1^2$ over all stabilizer decompositions of ψ. We find that ξ is easier to work with than the approximate stabilizer rank. For any fixed n-qubit state ψ, $\xi(\psi)$ can be computed using a simple convex optimization program, although the size of this computation scales poorly with n. In this section we develop tools that allow us to efficiently compute $\xi(\psi)$ whenever ψ is a tensor product of $1, 2$ and 3 qubit states. In particular, we prove Proposition 1 which establishes that ξ is multiplicative for tensor products of 1, 2, and 3-qubit states.

In subsection 6.1 we use standard convex duality to give a characterization of ξ in terms of the *stabilizer fidelity*, defined as the maximum overlap with respect to the set of stabilizer states

$$F(\psi) := \max_{\phi \in \mathrm{STAB}_n} |\langle \psi | \phi \rangle|^2. \tag{125}$$

As a consequence, multiplicativity of ξ is directly related to multiplicativity of the stabilizer fidelity. In subsection 6.2 we give sufficient and necessary conditions for multiplicativity of the stabilizer fidelity. In particular, we define the class of *stabilizer-aligned* states for which multiplicativity holds. In subsection 6.3 we investigate the class of stabilizer-aligned states and prove that all tensor products of $1, 2$ and 3 qubit states are stabilizer-aligned. Finally, in section 6.4 we use these results to prove Proposition 1.

6.1 Convex duality

Here we show that the optimization of $\xi(\psi)$ can be recast as a dual convex problem and we prove the following:

Theorem 4. *For any n-qubit state ψ we have*

$$\xi(\psi) = \max_{\omega} \frac{|\langle \psi | \omega \rangle|^2}{F(\omega)}, \tag{126}$$

where the maximum is over all n-qubit states ω.

Thus any n-qubit state ω can act as a witness to provide a lower bound on ξ and, furthermore, there exists at least one optimal witness state $\omega_\star$ which achieves the maximum in Eq. (126). For example, choosing $\omega = \psi$, we get the lower bound

$$\xi(\psi) \geq \frac{1}{F(\psi)}. \tag{127}$$

For Clifford magic states this lower bound is tight as stated in Proposition 2. We remark that Thm. 4 is a special case of results found in the literature on general resource theories [50].

Proof. We shall map the problem into the language of convex optimization and use standard results in that field [10]. Using the computation basis $\{|x\rangle\}$ we can decompose any stabilizer state $|\psi_j\rangle = \sum_x M_{x,j}|x\rangle$. Given a state $|\psi\rangle = \sum_x a_x |x\rangle$, the primal optimization problem can be written as

$$\sqrt{\xi(\psi)} = \min_c f(c) = ||c||_1 \tag{128}$$

$$\text{such that } Mc - a = 0 \tag{129}$$

This is clearly a convex optimization problem with affine constraints. Because the coefficient in c are complex, rather than real, this is a second order cone problem [10]. For any convex optimization problem there exists a dual function

$$g(\nu) = \inf_c \left(||c||_1 + \nu^T (Mc - a) \right) \tag{130}$$

$$= \begin{cases} -\nu^T a & \text{when } ||M^T \nu||_\infty \leq 1 \\ -\infty & \text{otherwise} \end{cases} \tag{131}$$

where for any value of the dual variables ν we have $g(\nu) \leq \sqrt{\xi(\psi)}$. The dual optimization problem is the maximisation of $g(\nu)$ over ν to obtain the best lower bound on $\sqrt{\xi(\psi)}$. We can discount the need for two cases by adding $||M^T \nu||_\infty \leq 1$ as a constraint, to obtain the problem

$$d^\star(\psi) = \max_\nu -\nu \cdot a \tag{132}$$
$$\text{such that } ||M^T \nu||_\infty \leq 1,$$

or more simply

$$d^\star(\psi) = \max_\nu \frac{-\nu \cdot a}{||M^T \nu||_\infty}. \tag{133}$$

Because the primal problem has affine constraints, we have strong duality and there must exist a $\nu_\star$ such that $g(\nu_\star) = -\nu_\star^T a = \sqrt{\xi(\psi)}$. Next, we restate this dual problem in terms of quantum states. For every ν we can associate a normalised quantum state

$$|\omega_\nu\rangle := \frac{1}{||\nu||_2} \sum_x (-\nu_x^*) |x\rangle, \tag{134}$$

so that

$$\langle \omega_\nu | \psi \rangle = \frac{-\nu \cdot a}{||\nu||_2}. \tag{135}$$

Next we note that

$$||M^T \nu||_\infty = \frac{\text{Max}_{|\phi\rangle \in \text{STAB}} |\langle \omega_\nu | \phi \rangle|}{||\nu||_2} = \frac{\sqrt{F(\omega_\nu)}}{||\nu||_2} \tag{136}$$

Therefore, the dual problem can also be stated as

$$d^\star(\psi) = \max_{|\omega_\nu\rangle} \frac{\langle \omega_\nu | \psi \rangle}{\sqrt{F(\omega_\nu)}}, \tag{137}$$

where the factors $||\nu||_2$ have cancelled out. The optimal $\nu_\star$ gives the optimal $|\omega_\star\rangle$, which completes the proof. □

6.2 Stabilizer alignment

Combining Theorems 2 and 1 we get an upper bound $\chi_\delta(\psi) \leq \delta^{-2} F(\psi)^{-1}$ on the approximate stabilizer rank of any Clifford magic state ψ. We shall be interested in the case when ψ is a tensor product of a large number of few-qubit magic states such as T-type or CCZ-type states. For example, the case $\psi = CCZ^{\otimes m}$ is relevant to gadget-based simulation of quantum circuits composed of Clifford gates and m CCZ gates. This motivates the question of whether the stabilizer fidelity $F(\psi)$ is multiplicative under tensor product, i.e.

$$F(\psi \otimes \phi) \stackrel{?}{=} F(\psi) F(\phi). \tag{138}$$

Note that $F(\psi \otimes \phi) \geq F(\psi) F(\phi)$ since the set of stabilizer states is closed under tensor product.

Below we define a set of quantum states $\mathcal{S}$ such that $F(\phi \otimes \psi) = F(\phi) F(\psi)$ whenever $\phi, \psi \in \mathcal{S}$. Remarkably, this set is also closed under tensor product, that is $\phi \otimes \psi \in \mathcal{S}$ whenever $\phi, \psi \in \mathcal{S}$. Moreover, we show that the stabilizer fidelity is not multiplicative for all states $\phi \notin \mathcal{S}$. More precisely, for any $\phi \notin \mathcal{S}$ there exists a state ψ such that $F(\phi \otimes \psi) > F(\phi) F(\psi)$. In that sense, our

results provide necessary and sufficient conditions under which the stabilizer fidelity is multiplicative under tensor product.

To state our results let us generalize the definition of stabilizer fidelity as follows. For each $n \geq 1$ and $0 \leq m \leq n$ define a set $S_{n,m}$ which consists of all stabilizer projectors Π on n qubits satisfying $\mathrm{Tr}[\Pi] = 2^m$.

Definition 9. *For any n-qubit state $|\phi\rangle$ define*

$$F_m(\phi) = 2^{-m/2} \max_{\Pi \in S_{n,m}} \langle \phi | \Pi | \phi \rangle. \qquad\qquad m = 0, \ldots, n.$$

Let us say that ϕ is stabilizer-aligned if $F_m(\phi) \leq F_0(\phi)$ for all m.

Note that in the above $F_0 = F$ is the stabilizer fidelity. Here we investigate the consequences of stabilizer-alignment. Whether or not a given state is stabilizer-aligned is discussed in the following subsection.

Theorem 5. *Suppose ϕ and ψ are stabilizer-aligned. Then $\phi \otimes \psi$ is stabilizer-aligned and*

$$F(\phi \otimes \psi) = F(\phi) F(\psi).$$

Conversely, suppose ϕ is not stabilizer-aligned. Let $\phi^\star$ be the complex conjugate of ϕ. Then

$$F(\phi \otimes \phi^\star) > F(\phi) F(\phi^\star).$$

The theorem implies that the stabilizer fidelity is multiplicative for any stabilizer-aligned states:

Corollary 1. *Suppose $\psi_1, \ldots, \psi_L$ are stabilizer-aligned quantum states. Then*

$$F(\psi_1 \otimes \psi_2 \otimes \ldots \otimes \psi_L) = \prod_{j=1}^{L} F(\psi_j).$$

We prove Theorem 5 using characterization of entanglement in tripartite stabilizer states from Ref. [13]:

Lemma 9 ([13])**.** *Any pure tripartite stabilizer state can be transformed by local unitary Clifford operators to a tensor product of states from the set $\{|0\rangle, |\Psi^+\rangle, |\Psi_3^+\rangle\}$ where*

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \qquad\qquad |\Psi_3^+\rangle = \frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right).$$

Corollary 2 ([13])**.** *Suppose Π be a stabilizer projector describing a bipartite system AB. Then there exists a unitary Clifford operator $U = U_A \otimes U_B$ and integers $a, b, c, d \geq 0$ such that*

$$U \Pi U^{-1} = \sum_{\alpha=1}^{2^a} \sum_{\beta=1}^{2^b} \sum_{\gamma=1}^{2^c} |\omega_{\alpha\beta\gamma}\rangle\langle\omega_{\alpha\beta\gamma}|, \tag{139}$$

where

$$|\omega_{\alpha\beta\gamma}\rangle = 2^{-d/2} \sum_{\delta=1}^{2^d} |\alpha, \gamma, \delta\rangle \otimes |\beta, \gamma, \delta\rangle. \tag{140}$$

Here $|\alpha, \gamma, \delta\rangle$ and $|\beta, \gamma, \delta\rangle$ are the computational basis vectors of A and B.

Proof. Let us apply Lemma 9 to a tripartite stabilizer state

$$|\Psi\rangle = (\Pi \otimes I) 2^{-n/2} \sum_{z\in\{0,1\}^n} |z\rangle_{AB} \otimes |z\rangle_C,$$

where $n = |A|+|B|$ and C is a system of n qubits. The lemma implies that Π is equivalent modulo local Clifford operators to a tensor product of local stabilizer projectors $|0\rangle\langle 0|$ and $I = |0\rangle\langle 0| + |1\rangle\langle 1|$ as well as bipartite projectors $|00\rangle\langle 00| + |11\rangle\langle 11|$ and $|\Psi^+\rangle\langle\Psi^+|$ shared between A and B. Let a and b be the number of times Π contains the identity factor on A and B respectively. Let c be the number of times Π contains the projector $|00\rangle\langle 00| + |11\rangle\langle 11|$ shared between A and B. Let d be the number of times Π contains the EPR projector $|\Psi^+\rangle\langle\Psi^+|$. The desired family of states $\omega_{\alpha\beta\gamma}$ is then obtained by writing each projector I and $|00\rangle\langle 00| + |11\rangle\langle 11|$ as a sum of rank-1 projectors onto the computational basis vectors. □

Proof of Theorem 5. To prove the first two claims of the theorem it suffices to show that

$$F_m(\phi\otimes\psi) \le F_0(\phi) F_0(\psi). \tag{141}$$

for all m. Indeed, combining Eq. (141) and the obvious bound $F_0(\phi)F_0(\psi) \le F_0(\phi\otimes\psi)$ shows that $F_m(\phi\otimes\psi) \le F_0(\phi\otimes\psi)$, that is, $\phi\otimes\psi$ is stabilizer-aligned. Using Eq. (141) for $m = 0$ gives multiplicativity of the stabilizer fidelity $F_0(\phi\otimes\psi) = F_0(\phi)F_0(\psi)$.

Define a bipartite system AB such that ϕ and ψ are states of A and B. Let Π be a stabilizer projector of rank 2^m such that

$$F_m(\phi\otimes\psi) = 2^{-m/2}\langle\phi\otimes\psi|\Pi|\phi\otimes\psi\rangle.$$

We shall write Π as a sum of rank-1 stabilizer projectors as stated in Corollary 2. Since local Clifford unitary operators do not change the stabilizer fidelity, we shall absorb the unitaries U_A and U_B into the states ϕ and ψ respectively. Accordingly, below we set $U = I$. Consider a single term $\omega_{\alpha\beta\gamma}$ in the decomposition of Π. Applying the Cauchy-Schwarz inequality one gets

$$|\langle\phi\otimes\psi|\omega_{\alpha\beta\gamma}\rangle|^2 = 2^{-d}\left|\sum_{\delta=1}^{2^d}\langle\phi|\alpha,\gamma,\delta\rangle\cdot\langle\psi|\beta,\gamma,\delta\rangle\right|^2 \le 2^{-d}\langle\phi|\Pi^A_{\alpha\gamma}|\phi\rangle\cdot\langle\psi|\Pi^B_{\beta\gamma}|\psi\rangle, \tag{142}$$

where we defined stabilizer projectors

$$\Pi^A_{\alpha,\gamma} = \sum_{\delta=1}^{2^d}|\alpha,\gamma,\delta\rangle\langle\alpha,\gamma,\delta| \quad\text{and}\quad \Pi^B_{\beta,\gamma} = \sum_{\delta=1}^{2^d}|\beta,\gamma,\delta\rangle\langle\beta,\gamma,\delta|. \tag{143}$$

By assumption, ψ is stabilizer-aligned. Thus

$$\max_\gamma \langle\psi|\sum_{\beta=1}^{2^b}\Pi^B_{\beta\gamma}|\psi\rangle \le 2^{(b+d)/2}F_0(\psi). \tag{144}$$

Here we noted that $\sum_{\beta=1}^{2^b}\Pi^B_{\beta\gamma}$ is a projector of rank 2^{b+d} for all γ. Combining Eq. (142,144) gives

$$\langle\phi\otimes\psi|\Pi|\phi\otimes\psi\rangle = \sum_{\alpha=1}^{2^a}\sum_{\beta=1}^{2^b}\sum_{\gamma=1}^{2^c}|\langle\phi\otimes\psi|\omega_{\alpha\beta\gamma}\rangle|^2 \le 2^{(b-d)/2}F_0(\psi)\cdot\langle\phi|\sum_{\alpha=1}^{2^a}\sum_{\gamma=1}^{2^c}\Pi^A_{\alpha,\gamma}|\phi\rangle \tag{145}$$

The assumption that ϕ is stabilizer-aligned gives

$$\langle\phi|\sum_{\alpha=1}^{2^a}\sum_{\gamma=1}^{2^c}\Pi^A_{\alpha,\gamma}|\phi\rangle \le 2^{(a+c+d)/2}F_0(\phi). \tag{146}$$

Here we noted that $\sum_{\alpha=1}^{2^a}\sum_{\gamma=1}^{2^c}\Pi^A_{\alpha,\gamma}$ is a projector of rank 2^{a+c+d}. Combining Eqs. (145,146) gives

$$\langle\phi\otimes\psi|\Pi|\phi\otimes\psi\rangle \le 2^{(a+b+c)/2}F_0(\psi)F_0(\phi).$$

It remains to notice that Π has rank 2^m, where $m = a+b+c$. This establishes Eq. (141).

We now prove the converse statement from Theorem 5.

Lemma 10. *Let ϕ be an n-qubit state which is not stabilizer-aligned. Then*

$$F_0(\phi \otimes \phi^\star) > F_0(\phi) F_0(\phi^\star).$$

Proof. If ϕ is not stabilizer-aligned then we have $F_m(\phi) > F_0(\phi)$ for some $m \in \{1, \ldots, n\}$. Let Π be a stabilizer projector with

$$F_m(\phi) = \frac{1}{\sqrt{2}^m} \langle \phi | \Pi | \phi \rangle.$$

Let C be an n-qubit Clifford such that

$$\Pi = C \left(|0\rangle\langle 0|_{n-m} \otimes I_m \right) C^\dagger.$$

Next consider a system of $2n$ qubits and partition them as $[2n] = ABA'B'$ where $|A| = |A'| = n-m$ and $|B| = |B'| = m$. Define a $2n$-qubit stabilizer state

$$|\theta\rangle = C \otimes \alpha |0\rangle_A |\Phi\rangle_{BB'} |0\rangle_{A'},$$

where

$$|\Phi\rangle_{BB'} = \frac{1}{\sqrt{2}^m} \sum_{z \in \{0,1\}^m} |z\rangle_B |z\rangle_{B'}.$$

Also define a normalized m-qubit state

$$|\omega\rangle = \frac{1}{2^{m/4} \sqrt{F_m(\phi)}} \left(\langle 0|_{n-m} \otimes I_m \right) C |\phi\rangle.$$

$$\begin{aligned} F_0(\phi \otimes \phi^\star) &\geq \langle \phi \otimes \phi^\star | \theta \rangle \langle \theta | \phi \otimes \phi^\star \rangle && (147) \\ &= \langle \omega \otimes \omega^\star | \Phi \rangle \langle \Phi | \omega \otimes \omega^\star \rangle 2^m (F_m(\phi))^2 && (148) \\ &= (F_m(\phi))^2 && (149) \\ &> F_0(\phi) F_0(\phi^\star). && (150) \end{aligned}$$

where in the last line we used the fact that $F_m(\phi) > F_0(\phi) = F_0(\phi^\star)$. □

□

6.3 Proving and disproving stabilizer alignment

In this section we prove that all states of $n \leq 3$ qubits are stabilizer-aligned. We also show that typical n-qubit states are not stabilizer-aligned for sufficiently large n. An important lemma is the following

Lemma 11. *For any quantum state ψ we have $F_m(\psi) \leq F_0(\psi)$ for $m = 1, 2, 3$.*

It follows immediately that

Corollary 3. *All states of $n \leq 3$ qubits are stabilizer-aligned.*

Indeed, if we consider n-qubit states, it suffices to check that $F_m(\psi) \leq F_0(\psi)$ for $m \leq n$.

Corollary 4. *If $F_0(\psi) \geq 1/4$ then ψ is stabilizer-aligned.*

Indeed, if $m \geq 4$ then $F_m(\psi) \leq 2^{-m/2} \leq 1/4 \leq F_0(\psi)$.

Finally, we show that Haar-random n-qubit states are not stabilizer-aligned for sufficiently large n.

Claim 2. *Let ψ be a Haar-random n-qubit state. Then*

$$\Pr[F_0(\psi \otimes \psi^\star) \neq F_0(\psi) F_0(\psi^\star)] \geq 1 - o(1).$$

and so for large enough n a typical state ψ is not stabilizer-aligned.

Highly structured states on a large number of qubits may be stabilizer-aligned, and for instance it is an open question whether or not all Clifford magic states are stabilizer-aligned.

Proof of Lemma 11. First, we claim that

$$F_{m-1}(\psi) \geq 2^{-1/2}\left(1+\left[\frac{2^m-1}{4^m-1}\right]^{1/2}\right)\cdot F_m(\psi) \tag{151}$$

for all $m \geq 1$. Indeed, consider a fixed m and a rank-2^m stabilizer projector $\Pi \in S_{n,m}$ such that $F_m(\psi) = 2^{-m/2}\langle\psi|\Pi|\psi\rangle$. Using the standard stabilizer formalism one can show that

$$U\Pi U^{-1} = I^{\otimes m}\otimes|0\rangle\langle 0|^{\otimes(n-m)} \equiv \Pi'$$

for some n-qubit unitary Clifford operator U. Define a state $|\psi'\rangle = U|\psi\rangle$. We have

$$\Pi'|\psi'\rangle = \Gamma^{1/2}|\omega\rangle\otimes|0^{n-m}\rangle$$

for some m-qubit normalized state $|\omega\rangle$ and $\Gamma = \langle\psi'|\Pi'|\psi'\rangle = \langle\psi|\Pi|\psi\rangle$. Since ω is normalized,

$$\sum_{P\neq I}\langle\omega|P|\omega\rangle^2 = 2^m-1,$$

where the sum runs over all 4^m-1 non-trivial Pauli operators on m qubits. Thus there exists an m-qubit Pauli operator $P\neq I$ such that

$$\langle\omega|P|\omega\rangle \geq \left(\frac{2^m-1}{4^m-1}\right)^{1/2}. \tag{152}$$

Define a stabilizer projector

$$\Pi'' = \frac{1}{2}(I+P)\otimes|0\rangle\langle 0|^{\otimes(n-m)} \in S_{n,m-1}.$$

Recalling that $\Gamma = \langle\psi|\Pi|\psi\rangle = 2^{m/2}F_m(\psi)$ we arrive at

$$\begin{aligned} F_{m-1}(\psi) = F_{m-1}(\psi') &\geq 2^{-(m-1)/2}\langle\psi'|\Pi''|\psi'\rangle \\ &= 2^{-(m-1)/2}\frac{\Gamma}{2}(1+\langle\omega|P|\omega\rangle) \\ &= 2^{-1/2}(1+\langle\omega|P|\omega\rangle)\cdot F_m(\psi). \end{aligned} \tag{153}$$

Combining this identity and Eq. (152) proves Eq. (151). Applying Eq. (151) inductively gives

$$F_0(\psi) \geq 2^{-1/2}(1+\sqrt{1/3})\cdot F_1(\psi) \approx 1.115\cdot F_1(\psi), \tag{154}$$

$$F_0(\psi) \geq 2^{-1/2}(1+\sqrt{1/3})\cdot 2^{-1/2}(1+\sqrt{3/15})\cdot F_2(\psi) \approx 1.141\cdot F_2(\psi), \tag{155}$$

$$F_0(\psi) \geq 2^{-1/2}(1+\sqrt{1/3})\cdot 2^{-1/2}(1+\sqrt{3/15})\cdot 2^{-1/2}(1+\sqrt{7/63})\cdot F_3(\psi) \approx 1.076\cdot F_3(\psi). \tag{156}$$

Thus $F_0(\psi) \geq F_m(\psi)$ for $m = 1,2,3$ proving the lemma. □

Next, we prove claim 2.

Proof. Let w be any n-qubit state. For Haar-random ψ the probability density function $p(y)$ of $y = |\langle w|\psi\rangle|^2$ does not depend on w and is equal to (equation (9) of Ref. [60]),

$$p(y) = (2^n-1)(1-y)^{2^n-2}.$$

Integrating this we obtain the cumulative distribution function

$$\Pr\left[|\langle w|\psi\rangle|^2 \geq x\right] = (1-x)^{2^n-1} \leq \exp(-x(2^n-1)).$$

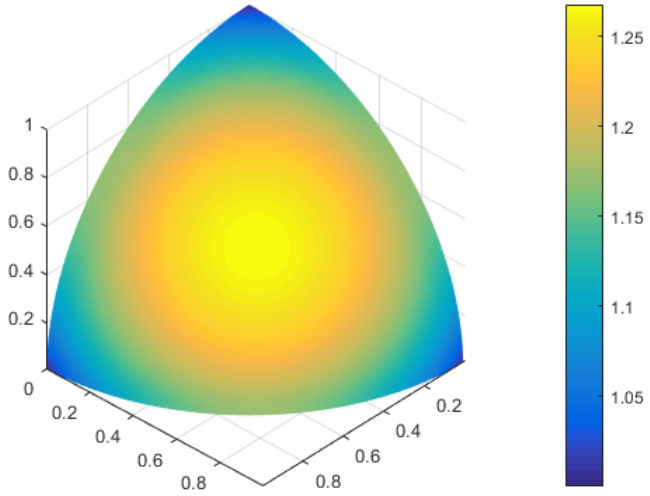


Figure 7: The color indicates the value of ξ for single-qubit states in the first octant of the Bloch sphere. This function controls the upper bound on the approximate stabilizer rank as in Eq. (162).

Since an n-qubit stabilizer state is specified by $O(n^2)$ bits the cardinality of the set STAB_n of n-qubit stabilizer states is $|\mathrm{STAB}_n| \leq 2^{O(n^2)}$. Choosing $x = n^3/2^n$ and applying a union bound we get

$$\Pr\left[\left(\max_{w\in\mathrm{STAB}_n} |\langle\psi|w\rangle|^2\right) \geq n^3/2^n\right] \leq e^{-\Omega(n^3)}.$$

This says that with probability very close to 1 a random ψ has $F_0(\psi) = F_0(\psi^\star) \leq n^3/2^n$. Next suppose ψ has this property. Then

$$F_0(\psi\otimes\psi^\star) \geq \left|\frac{1}{\sqrt{2}^n}\sum_{z\in\{0,1\}^n}\langle z|\psi\rangle\langle z|\psi^\star\rangle\right|^2 = \frac{1}{2^n},$$

which is strictly greater than $F_0(\psi)F_0(\psi^\star) \leq 2^{-2n}(n^3)^2$. □

6.4 Multiplicativity of stabilizer extent

This subsection considers tensor products of few-qubit states that involve at most three qubits each and shows that ξ behaves multiplicatively for such products, proving Proposition 1. The proof will draw heavily on Theorem 4 and Corollary 3.

Proof of Proposition 1. By Theorem 4 there exist witness states $\{\omega_{\star,1}, \omega_{\star,2}, \ldots, \omega_{\star,L}\}$ such that

$$\frac{|\langle\psi_j|\omega_{\star,j}\rangle|^2}{F(\omega_{\star,j})} = \xi(\psi_j). \tag{157}$$

We consider the product witness $|\Omega\rangle = \bigotimes_j |\omega_{\star,j}\rangle$ for which

$$|\langle\Psi|\Omega\rangle|^2 = \prod_j |\langle\psi_j|\omega_{\star,j}\rangle|^2. \tag{158}$$

Furthermore, using Corollary 3 and Theorem 5 we get

$$F(\Omega) = \prod_j F(\omega_{\star,j}). \tag{159}$$

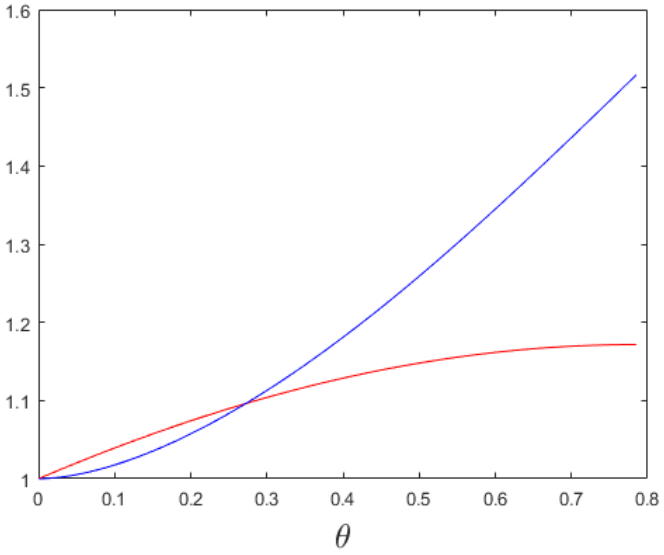


Figure 8: The approximate stabilizer rank of $|\theta^{\otimes n}\rangle$ is upper bounded as $\chi_\delta(\theta^{\otimes n}) \leq \delta^{-2}\xi(\theta)^n$, where $\xi(\theta) = (\cos(\theta/2) + \tan(\pi/8)\sin(\theta/2))^2$ is attained by the stabilizer decomposition from Eq. (163). The red line shows the function $\xi(\theta)$ for $\theta \in [0, \pi/4]$ and the blue line shows the function $g(\theta) = 2^{h_2(\cos^2(\theta/2))}$ where h_2 is the binary entropy. Our upper bound on the approximate stabilizer rank of $\theta^{\otimes n}$ performs better that obtained by a naive expansion in the $0, 1$ basis whenever the red line lies below the blue line.

Putting this together yields

$$\frac{|\langle\Psi|\Omega\rangle|^2}{F(\Omega)} = \prod_j \frac{|\langle\psi_j|\omega_{\star,j}\rangle|^2}{F(\omega_{\star,j})} = \prod_{j=1}^{L} \xi(\psi_j). \tag{160}$$

Thus, using Ω as a witness, we get

$$\prod_{j=1}^{L} \xi(\psi_j) \leq \xi(\Psi). \tag{161}$$

Furthermore, ξ is inherently sub-multiplicative and so we must have equality. □

Now let us see how this can be used to bound the approximate stabilizer rank of a product state $\alpha^{\otimes n}$ where α is a single-qubit state. Combining Theorem 1 with Lemma 6 we get

$$\chi_\delta(\alpha^{\otimes n}) \leq \delta^{-1}\xi(\alpha^{\otimes n}) = \delta^{-2}(\xi(\alpha))^n. \tag{162}$$

Note that since α is a single-qubit state we can easily compute $\xi(\alpha)$ by solving a small convex optimization program. In Figure 7 we plot $\xi(\alpha)$ as a function of the single-qubit state α on the first octant of the Bloch sphere.

The maximum value plotted in Figure 7 is $\xi(f) = 2/(1 + 1/\sqrt{3}) \approx 1.2679$, which is achieved by the so-called face state $|f\rangle$ which lies in the center of the surface and is defined by

$$|f\rangle\langle f| = \frac{1}{2}\left(I + \frac{1}{\sqrt{3}}(X + Y + Z)\right).$$

The single-qubit states in Figure 7 which lie in the x-z plane are of the form

$$|\theta\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle = (\cos(\theta/2) - \sin(\theta/2))\,|0\rangle + \sqrt{2}\sin(\theta/2)|+\rangle \tag{163}$$

for $\theta \in [0, \pi/2]$. In this case, the stabilizer decomposition on the right hand side achieves the optimal value of ξ. We can use this example to show that in the general case the upper bound on approximate stabilizer rank given in Theorem 1 is not tight (for $\delta = O(1)$, say). When θ is close

to 0 it becomes advantageous to expand $\theta^{\otimes n}$ in the standard $0, 1$ basis and truncate amplitudes which are very small. Using this approach one obtains an approximate stabilizer rank scaling as $2^{h_2(\cos^2(\theta/2))}$ where h_2 is the binary entropy. In Figure 8 we compare the performance of these upper bounds as a function of θ.

7 Acknowledgements

EC and MH are supported by the EPSRC (Grant No. EP/M024261/1). PC is supported by the EPSRC (Grant No. EP/L015242/1). The collaboration benefited from support by the NQIT project partnership fund (Grant No. EP/M013243/1), an EPSRC IIKE award and the IBM Research Frontiers Institute.

References

[1] Scott Aaronson and Lijie Chen. Complexity-theoretic foundations of quantum supremacy experiments. In *32nd Computational Complexity Conference (CCC 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. DOI: 10.4230/LIPIcs.CCC.2017.22.

[2] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004. DOI: 10.1103/PhysRevA.70.052328.

[3] Dorit Aharonov, Michael Ben-Or, Elad Eban, and Urmila Mahadev. Interactive proofs for quantum computations. *arXiv preprint arXiv:1704.04487*, 2017.

[4] Gadi Aleksandrowicz, Thomas Alexander, Panagiotis Barkoutsos, Luciano Bello, Yael Ben-Haim, David Bucher, Francisco Jose Cabrera-Hernández, Jorge Carballo-Franquis, Adrian Chen, Chun-Fu Chen, Jerry M. Chow, Antonio D. Córcoles-Gonzales, Abigail J. Cross, Andrew Cross, Juan Cruz-Benito, Chris Culver, Salvador De La Puente González, Enrique De La Torre, Delton Ding, Eugene Dumitrescu, Ivan Duran, Pieter Eendebak, Mark Everitt, Ismael Faro Sertage, Albert Frisch, Andreas Fuhrer, Jay Gambetta, Borja Godoy Gago, Juan Gomez-Mosquera, Donny Greenberg, Ikko Hamamura, Vojtech Havlicek, Joe Hellmers, Łukasz Herok, Hiroshi Horii, Shaohan Hu, Takashi Imamichi, Toshinari Itoko, Ali Javadi-Abhari, Naoki Kanazawa, Anton Karazeev, Kevin Krsulich, Peng Liu, Yang Luh, Yunho Maeng, Manoel Marques, Francisco Jose Martín-Fernández, Douglas T. McClure, David McKay, Srujan Meesala, Antonio Mezzacapo, Nikolaj Moll, Diego Moreda Rodríguez, Giacomo Nannicini, Paul Nation, Pauline Ollitrault, Lee James O'Riordan, Hanhee Paik, Jesús Pérez, Anna Phan, Marco Pistoia, Viktor Prutyanov, Max Reuter, Julia Rice, Abdón Rodríguez Davila, Raymond Harry Putra Rudy, Mingi Ryu, Ninad Sathaye, Chris Schnabel, Eddie Schoute, Kanav Setia, Yunong Shi, Adenilton Silva, Yukio Siraichi, Seyon Sivarajah, John A. Smolin, Mathias Soeken, Hitomi Takahashi, Ivano Tavernelli, Charles Taylor, Pete Taylour, Kenso Trabing, Matthew Treinish, Wes Turner, Desiree Vogt-Lee, Christophe Vuillot, Jonathan A. Wildstrom, Jessica Wilson, Erick Winston, Christopher Wood, Stephen Wood, Stefan Wörner, Ismail Yunus Akhalwaya, and Christa Zoufal. Qiskit: An open-source framework for quantum computing, 2019.

[5] Noga Alon. Transversal numbers of uniform hypergraphs. *Graphs and Combinatorics*, 6(1): 1–4, 1990. DOI: 10.1007/BF01787474.

[6] Simon Anders and Hans J Briegel. Fast simulation of stabilizer circuits using a graph-state representation. *Physical Review A*, 73(2):022334, 2006. DOI: 10.1103/PhysRevA.73.022334.

[7] Ryan S. Bennink, Erik M. Ferragut, Travis S. Humble, Jason A. Laska, James J. Nutaro, Mark G. Pleszkoch, and Raphael C. Pooser. Unbiased simulation of near-Clifford quantum circuits. *Physical Review A*, 95:062337, Jun 2017. DOI: 10.1103/PhysRevA.95.062337.

[8] Sergio Boixo, Sergei V Isakov, Vadim N Smelyanskiy, and Hartmut Neven. Simulation of low-depth quantum circuits as complex undirected graphical models. *arXiv preprint arXiv:1712.05384*, 2017.

[9] Sergio Boixo, Sergei V Isakov, Vadim N Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J Bremner, John M Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14(6):595, 2018. DOI: 10.1038/s41567-018-0124-x.

[10] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[11] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by Clifford gates. *Physical Review Letters*, 116(25):250501, 2016. DOI: 10.1103/PhysRevLett.116.250501.

[12] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316, 2005. DOI: 0.1103/PhysRevA.71.022316.

[13] Sergey Bravyi, David Fattal, and Daniel Gottesman. Ghz extraction yield for multipartite stabilizer states. *Journal of Mathematical Physics*, 47(6):062106, 2006. DOI: 10.1063/1.2203431.

[14] Sergey Bravyi, Graeme Smith, and John A. Smolin. Trading classical and quantum computational resources. *Physical Review X*, 6:021043, Jun 2016. DOI: 10.1103/PhysRevX.6.021043.

[15] Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Physical Review Letters*, 117 (8):080501, 2016. DOI: 0.1103/PhysRevLett.117.080501.

[16] Earl T. Campbell. Catalysis and activation of magic states in fault-tolerant architectures. *Physical Review A*, 83:032317, Mar 2011. DOI: 10.1103/PhysRevA.83.032317.

[17] Jianxin Chen, Fang Zhang, Mingcheng Chen, Cupjin Huang, Michael Newman, and Yaoyun Shi. Classical simulation of intermediate-size quantum circuits. *arXiv preprint arXiv:1805.01450*, 2018.

[18] Elizabeth Crosson and John Bowen. Quantum ground state isoperimetric inequalities for the energy spectrum of local hamiltonians. *arXiv preprint arXiv:1703.10133*, 2017.

[19] Koen De Raedt, Kristel Michielsen, Hans De Raedt, Binh Trieu, Guido Arnold, Marcus Richter, Th Lippert, H Watanabe, and N Ito. Massively parallel quantum computer simulator. *Computer Physics Communications*, 176(2):121–136, 2007. DOI: 10.1016/j.cpc.2006.08.007.

[20] Nicolas Delfosse, Philippe Allard Guerin, Jacob Bian, and Robert Raussendorf. Wigner function negativity and contextuality in quantum computation on rebits. *Physical Review X*, 5: 021003, Apr 2015. DOI: 10.1103/PhysRevX.5.021003.

[21] Lior Eldar and Aram W Harrow. Local Hamiltonians whose ground states are hard to approximate. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 427–438. IEEE, 2017. DOI: 10.1109/FOCS.2017.46.

[22] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem. *arXiv preprint arXiv:1412.6062*, 2014.

[23] Austin G Fowler, Simon J Devitt, and Cody Jones. Surface code implementation of block code state distillation. *Scientific reports*, 3:1939, 2013. DOI: 10.1038/srep01939.

[24] E Schuyler Fried, Nicolas PD Sawaya, Yudong Cao, Ian D Kivlichan, Jhonathan Romero, and Alán Aspuru-Guzik. qtorch: The quantum tensor contraction handler. *PloS one*, 13(12): e0208510, 2018. DOI: 10.1371/journal.pone.0208510.

[25] Hector J Garcia, Igor L Markov, and Andrew W Cross. Efficient inner-product algorithm for stabilizer states. *arXiv preprint arXiv:1210.6646*, 2012.

[26] Héctor J. García, Igor L. Markov, and Andrew W. Cross. On the geometry of stabilizer states. *Quantum Information & Computation*, 14:683, 2014.

[27] Daniel Gottesman. Theory of fault-tolerant quantum computation. *Physical Review A*, 57(1): 127, 1998. DOI: 10.1103/PhysRevA.57.127.

[28] Daniel Gottesman and Isaac L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402:390, 1999. DOI: 10.1038/46503.

[29] David Gross, Sepehr Nezami, and Michael Walter. Schur-Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations. *arXiv preprint arXiv:1712.08628*, 2017.

[30] Thomas Häner and Damian S Steiger. 0.5 petabyte simulation of a 45-qubit quantum circuit. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, page 33. ACM, 2017. DOI: 10.1145/3126908.3126947.

[31] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[32] Mark Howard and Earl Campbell. Application of a resource theory for magic states to fault-tolerant quantum computing. *Physical Review Letters*, 118:090501, Mar 2017. DOI: 10.1103/PhysRevLett.118.090501.
[33] Cupjin Huang, Michael Newman, and Mario Szegedy. Explicit lower bounds on strong quantum simulation. *arXiv preprint arXiv:1804.10368*, 2018.
[34] Cody Jones. Low-overhead constructions for the fault-tolerant Toffoli gate. *Physical Review A*, 87(2):022328, 2013. DOI: 10.1103/PhysRevA.87.022328.
[35] Richard Jozsa and Sergii Strelchuk. Efficient classical verification of quantum computations. *arXiv preprint arXiv:1705.02817*, 2017.
[36] Angela Karanjai, Joel J Wallman, and Stephen D Bartlett. Contextuality bounds the efficiency of classical simulation of quantum processes. *arXiv preprint arXiv:1802.07744*, 2018.
[37] Lucas Kocia and Peter Love. Discrete Wigner formalism for qubits and noncontextuality of Clifford gates on qubit stabilizer states. *Physical Review A*, 96(6):062134, 2017. DOI: 10.1103/PhysRevA.96.062134.
[38] Richard Kueng and David Gross. Qubit stabilizer states are complex projective 3-designs. *arXiv preprint arXiv:1510.02767*, 2015.
[39] Riling Li, Bujiao Wu, Mingsheng Ying, Xiaoming Sun, and Guangwen Yang. Quantum supremacy circuit simulation on Sunway TaihuLight. *arXiv preprint arXiv:1804.04797*, 2018.
[40] Igor L Markov and Yaoyun Shi. Simulating quantum computation by contracting tensor networks. *SIAM Journal on Computing*, 38(3):963–981, 2008. DOI: 10.1137/050644756.
[41] Servet Martínez, Gérard Michon, and Jaime San Martín. Inverse of strictly ultrametric matrices are of Stieltjes type. *SIAM Journal on Matrix Analysis and Applications*, 15(1):98–106, 1994. DOI: 10.1137/S0895479891217011.
[42] Dmitri Maslov and Martin Roetteler. Shorter stabilizer circuits via Bruhat decomposition and quantum circuit transformations. *arXiv preprint arXiv:1705.09176*, 2017.
[43] David C McKay, Christopher J Wood, Sarah Sheldon, Jerry M Chow, and Jay M Gambetta. Efficient Z gates for quantum computing. *Physical Review A*, 96(2):022330, 2017. DOI: 10.1103/PhysRevA.96.022330.
[44] Tomoyuki Morimae and Joseph F Fitzsimons. Post hoc verification with a single prover. *Physical Review Letters*, 120:040501, 2018. DOI: 10.1103/PhysRevLett.120.040501.
[45] Reinhard Nabben and Richard S Varga. A linear algebra proof that the inverse of a strictly ultrametric matrix is a strictly diagonally dominant Stieltjes matrix. *SIAM Journal on Matrix Analysis and Applications*, 15(1):107–113, 1994. DOI: 10.1137/S0895479892228237.
[46] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.
[47] Hakop Pashayan, Joel J Wallman, and Stephen D Bartlett. Estimating outcome probabilities of quantum circuits using quasiprobabilities. *Physical Review Letters*, 115(7):070501, 2015. DOI: 10.1103/PhysRevLett.115.070501.
[48] Edwin Pednault, John A Gunnels, Giacomo Nannicini, Lior Horesh, Thomas Magerlein, Edgar Solomonik, and Robert Wisnieff. Breaking the 49-qubit barrier in the simulation of quantum circuits. *arXiv preprint arXiv:1710.05867*, 2017.
[49] John Preskill. Quantum computing in the NISQ era and beyond. *arXiv preprint arXiv:1801.00862*, 2018.
[50] Bartosz Regula. Convex geometry of quantum resource quantification. *Journal of Physics A: Mathematical and Theoretical*, 51(4):045303, 2017. DOI: 10.1088/1751-8121/aa9100.
[51] M. Rötteler. Quantum algorithms for highly non-linear Boolean functions. In *Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms*, pages 448–457, 2010.
[52] Mikhail Smelyanskiy, Nicolas PD Sawaya, and Alán Aspuru-Guzik. qHiPSTER: the quantum high performance software testing environment. *arXiv preprint arXiv:1601.07195*, 2016.
[53] W. van Dam, S. Hallgren, and L. Ip. Quantum Algorithms for Some Hidden Shift Problems. *SIAM Journal on Computing*, 36(3):763–778, January 2006. ISSN 0097-5397.
[54] Maarten Van Den Nest. Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond. *Quantum Information & Computation*, 10(3):258–271, 2010.
[55] Maarten Van den Nest. Simulating quantum computers with probabilistic methods. *Quantum Information & Computation*, 11(9-10):784–812, 2011.

[56] Victor Veitch, Christopher Ferrie, David Gross, and Joseph Emerson. Negative quasi-probability as a resource for quantum computation. *New Journal of Physics*, 14(11):113011, 2012. DOI: 10.1088/1367-2630/14/11/113011.

[57] Zak Webb. The clifford group forms a unitary 3-design. *Quantum Information & Computaion*, 16:1379, 2016.

[58] Ulli Wolff, Alpha Collaboration, et al. Monte Carlo errors with less errors. *Computer Physics Communications*, 156(2):143–153, 2004. DOI: 10.1016/S0010-4655(03)00467-3.

[59] Huangjun Zhu, Richard Kueng, Markus Grassl, and David Gross. The Clifford group fails gracefully to be a unitary 4-design. *arXiv preprint arXiv:1609.08172*, 2016.

[60] Karol Zyczkowski and Hans-Jürgen Sommers. Truncations of random unitary matrices. *Journal of Physics A: Mathematical and General*, 33(10):2045, 2000. DOI: 10.1088/0305-4470/33/10/307.