eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Application of a resource theory for magic states to fault-tolerant quantum computing

Mark Howard[*] and Earl Campbell[†]

Department of Physics and Astronomy, University of Sheffield, Sheffield, UK

Motivated by their necessity for most fault-tolerant quantum computation schemes, we formulate a resource theory for magic states. We first show that robustness of magic is a well-behaved magic monotone that operationally quantifies the classical simulation overhead for a Gottesman-Knill type scheme using ancillary magic states. Our framework subsequently finds immediate application in the task of synthesizing non-Clifford gates using magic states. When magic states are interspersed with Clifford gates, Pauli measurements and stabilizer ancillas—the most general synthesis scenario— then the class of synthesizable unitaries is hard to characterize. Our techniques can place non-trivial lower bounds on the number of magic states required for implementing a given target unitary. Guided by these results we have found new and optimal examples of such synthesis.

Quantum resource theories attempt to capture what is quintessentially quantum in a piece of technology. For example, entanglement is the relevant resource for quantum cryptography and communication. The resource framework for entanglement finds practical application in bounding the efficiency of entanglement distillation protocols. An abundance of other resource theories have been related to various aspects of quantum theory [1–8]. Once a quantum computer is made fault-tolerant, some computational operations become relatively easy, and some more difficult, leading to a natural resource picture called the magic state model [9, 10] (although alternative routes to fault-tolerant universality exist [11]). Preparation of stabilizer states and implementation of Clifford unitaries and Pauli measurements constitute free resources. Difficult operations include preparation of magic states, a supply of which is necessary in order to promote the easier operations to a universal gate set. With only free resources, the computation can be efficiently classically simulated, whereas with a liberal supply of pure magic states, universal quantum computation is unlocked. For qudit (d-level) quantum computers with odd d, a resource theory of magic (or equivalently contextuality with respect to stabilizer measurements [12, 13]) has been developed [7, 14, 15]. This relies on a well-behaved discrete Wigner function [16], which in turn relies on quirks of odd dimensional Hilbert space. Here we address the most practically important case by quantifying the magic for multiqubit systems, relating this resource measure to simulation complexity and applying the resource theory to the practical problem of gate synthesis.

The canonical magic state is $|H\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$, which enables application of a single-qubit unitary $T = \text{diag}(1, e^{i\pi/4})$ [9, 10]. A circuit composed of elements from the Clifford+T gate set acting on the standard computational basis input suffices for universal quantum computation. Such a circuit can be classically simulated, but in a time that scales exponentially in the number of T gates [17]. Faster simulation algorithms were recently discovered that relate the simulation complexity to the stabilizer rank [18–20], a measure of magic for pure

states. Such techniques do not naturally adapt to mixed magic states, and stabilizer rank is qualitatively very different to the magic measure we establish here. For quantum computations using qudits with odd dimension, the discrete Wigner function provides a quasiprobabilitiy distribution and Pashayan et al. [21] showed that the negativity quantifies the simulation complexity. Here we provide a general simulation scheme, which can be naturally applied to mixed-state qubit quantum computations using any kind of ancillary magic state (e.g., a multi-qubit magic state enabling a Toffoli gate). Furthermore, for many problems our approach is competitive with comparable schemes based on stabilizer rank [18, 19].

Due to the high price assigned to $|H\rangle$ states and hence T gates, it behooves us to find Clifford+T circuit implementations of quantum algorithms that are parsimonious in their use of T gates. The topic of circuit synthesis has made tremendous progress in recent years [22–30], compared with Solovay-Kitaev type constructions that were for a long time the standard benchmark. Developments include identifying special algebraic forms for all gates that can be unitarily synthesized over the Clifford+T gate set [24] or over the smaller CNOT+T gate set [22, 23]. However, circuit synthesis need not be a purely unitary process, and more generally may be aided by ancillary stabilizer states, measurements and classical feed-forward. There are hints that general synthesis can be significantly more powerful [26, 27, 30, 31], though the paradigm is not well understood. Our resource framework helps with this problem by establishing non-trivial lower bounds on the number of $|H\rangle$ states, or equivalently T gates, required for the general synthesis scenario. This allows us to identify several circuits as optimal. Such resource-theoretic tools work for any kind of state, not just $|H\rangle$, but they are particularly well-motivated for magic states from the third level of the Clifford hierarchy e.g., Toffoli resource states. We note how general synthesis has a curious relationship to Clifford equivalence of magic states. From this vantage point, we discover several new examples of general synthesis protocols with resource savings over previous unitary synthesis methods.

The Supplementary Materials (text, which includes Refs. [32–36], and files) contain numerous additional computations. Results include identifying the most robust states, most robust gates and classification of all three and four qubit diagonal gates from the third level of the Clifford hierarchy. Intriguingly, one maximally robust state is the Hoggar [37] fiducial state.

Robustness of Magic.—Vidal and Tarrach [8] showed that the amount of separable noise that makes an entangled state become separable is an entanglement monotone, which they called robustness. The basic principle can been adapted for use in other resource theories with a set of free states. Denoting $\mathcal{S}_n = \{\sigma_i\}$ as the set of pure n-qubit stabilizer states, we define the robustness of magic as

$$\text{(RoM)} \qquad \mathcal{R}(\rho) = \min_x \left\{ \sum_i |x_i| ; \rho = \sum_i x_i \sigma_i \right\} . \qquad (1)$$

Decompositions of the form $\sum_i x_i \sigma_i$ are called stabilizer pseudomixtures. We have $\sum_i x_i = 1$, but x_i may be negative and so they provide a quasiprobability distribution. The optimization in (1) can be rewritten in terms of a linear system as

$$\mathcal{R}(\rho) = \min ||x||_1 \text{ subject to } Ax = b, \qquad (2)$$

where $||x||_1 = \sum_i |x_i|$, $b_i = \text{Tr}(P_i \rho)$ and $A_{j,i} = \text{Tr}(P_j \sigma_i)$ where P_j is the jth Pauli operator. For example, consider the single-qubit magic state $|H\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$, then in the Pauli operator basis

$$A = \begin{array}{c} \langle \mathbb{1} \rangle \\ \langle X \rangle \\ \langle Y \rangle \\ \langle Z \rangle \end{array} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix},$$

and the solution of (2) is $x = \left(\sqrt{2}, 0, 1, 1 - \sqrt{2}, 0, 0 \right)/2$ implying $\mathcal{R}(|H\rangle) = \sqrt{2}$. There are a number of freely available solvers for linear programs [38, 39], which are efficient in the size of A. From our formulation of the problem it is clear that $\min_{Ax=b} ||x||_1$ is feasible and bounded. Consequently, strong duality holds i.e.,

$$\min_{Ax=b} ||x||_1 = \max_{||A^T y||_\infty \le 1} -b^T y, \qquad (3)$$

and the aforementioned solvers can provide a certificate y of optimality [40]. Despite the theoretical efficiency of the linear programming problem, the number of stabilizer states in $\mathcal{P}_n$ scales super-exponentially with n, so that $|\mathcal{S}_n| = 2^n \prod_{j=1}^n (2^j + 1)$ [16]. Practically we are limited to $1 \le n \le 5$ qubits. We have made available the corresponding A matrices in Supplementary Material.

Robustness of Magic (RoM) possesses all the desirable qualities of a resource theoretic measure (see Supplementary Material for proofs of the following). For a
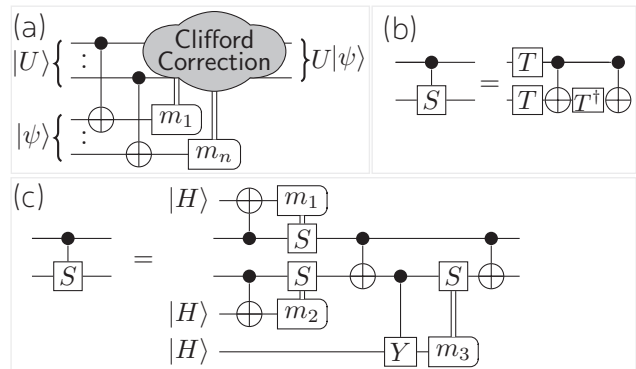

FIG. 1. (a) Half-teleportation gadget for implementing diagonal U in the third level of the Clifford hierarchy. The circuit uses a resource state $|U\rangle = U|+\rangle^{\otimes n}$. (b) Exact synthesis of CS gate using T-gates and CNOTS. (c) The CS circuit as a gadgetized circuit using three $|H\rangle$ magic states. Our techniques show this synthesis is provably optimal.

mixed stabilizer state, we find $x_i > 0$ entailing $\sum_i |x_i| = \sum_i x_i = 1$. For a non-stabilizer state, at least one x_i is negative and then RoM must exceed unity. Therefore RoM is faithful. Crucially, RoM is non-increasing under stabilizer operations, the free set of operations in the resource theory. Finally, RoM is submultiplicative,

$$\mathcal{R}(\rho_1 \otimes \rho_2) \le \mathcal{R}(\rho_1)\mathcal{R}(\rho_2), \qquad (4)$$

which follows by using the minimal stabilizer pseudomixtures for ρ_1 and ρ_2 to construct a not-necessarily-minimal stabilizer pseudomixture for $\rho_1 \otimes \rho_2$. Useful lower bounds on $\mathcal{R}(\rho_1 \otimes \rho_2)$ can also be obtained (see Supplementary Material).

Resource theoretic frameworks are commonplace in quantum information theory, but do not always directly lend themselves to an operational meaningful interpretation or to useful applications in solving relevant problems. In the next section we show how RoM quantifies the exponential simulation overhead for a version of the Gottesman-Knill protocol where non-stabilizer ancillas ρ can be added to an otherwise stabilizer circuit. The subsequent section discusses RoM's application to the task of implementing non-Clifford operations in an economical way.

Robustness quantifies classical simulation overhead.— The Gottesman-Knill (GK) theorem shows that for any stabilizer circuit written as a superoperator $\mathcal{E}$ and pure stabilizer state σ_i, we can efficiently sample from the outcome of a Pauli measurement on $\mathcal{E}(\sigma_i)$. By collecting many samples, we can estimate the expectation value to any desired accuracy. If the input state is a probabilistic ensemble of stabilizer states, the Gottesman-Knill theorem still holds provided we can efficiently sample from the ensemble. A Clifford+T circuit suffering very heavy noise can be simulated in this way [41–43]. Here we provide a simulation algorithm for estimating

an expectation value $P_\rho := \text{tr}[P\mathcal{E}(\rho)]$ where $\mathcal{E}$ is a stabilizer operation. The simulation time cost scales with $\sum_i |x_i| \geq \mathcal{R}(\rho)$ where x_i are the quasiprobabilities used (which may be suboptimal). First, as in [21], we use the quasiprobability distribution x_i to form the probability distribution $p_i = |x_i|/\sum_i |x_i|$. We sample an i value from this probability distribution and use GK to obtain an eigenvalue $m = \pm 1$ for the Pauli measurement on $\mathcal{E}(\sigma_i)$. Our simulation outputs not m, but $M = \text{sign}(x_i)m\sum_i |x_i|$ where $\text{sign}(x_i) = 1$ for $x_i > 0$ and $\text{sign}(x_i) = -1$ otherwise. Notice that each run outputs $\pm\sum_i |x_i|$ and not ± 1, which leads to a larger variance of our random variable. We repeat this sampling process many times and find the mean value of M, which gives an unbiased estimator of P_ρ. The Hoeffding inequalities show that for random variables bounded in the interval $[-\sum_i |x_i|, +\sum_i |x_i|]$, N samples will estimate the mean to within δ of the actual mean with probability exceeding $1 - \epsilon$ where $\epsilon = 2\exp(-N\delta^2/2(\sum_i |x_i|)^2)$. In other words, the desired accuracy is guaranteed by using

$$N = \frac{2}{\delta^2}(\sum_i |x_i|)^2 \ln\left(\frac{2}{\epsilon}\right) \qquad (5)$$

samples. Using an optimal stabilizer pseudomixture, the number of samples scales quadratically in the robustness, though the robustness typically scales exponentially in the number of magic states. For each of these samples, the GK scheme requires a polynomial amount of time provided we know how to efficiently sample from the quasiprobabilitiy distribution.

Nonstabilizer circuits.— For any quantum circuit we can find an equivalent gadgetized version [9, 18, 19] over the Clifford plus T gate set; all uses of T are replaced with the standard state injection circuit whereby a $|H\rangle$ state is entangled with a data qubit and subsequently measured out (see Fig. 1 for an example). The T gadget is just one example from an infinitely large family of similar gadgets. All diagonal gates from the third level of the Clifford hierarchy—the set of gates that map Pauli operators to Clifford gates under conjugation—are also suitable for gadgetization. These gates are sufficient to promote Clifford gates to universality, and have the added property that access to the state $|U\rangle = U|+\rangle^{\otimes k}$ allows for deterministic implementation of the gate U [44, 45], as depicted in Fig. 1. This family includes important multiqubit gates such as the control-control-Z (CCZ), which is Clifford equivalent to the Toffoli, and control-S (CS) where $S = T^2$. Therefore, a quantum circuit $C_1 U_1 C_2 U_2 \ldots U_N C_{N+1}$ where C_i are Clifford is equivalent to a stabilizer circuit consuming the resource $|U\rangle$ where $U := \otimes U_j$. We remark that diagonal, third-level gates are exactly synthesizable from CNOT and T gates [22, 23, 46, 47].

Large resource states.— For large resource states the exact robustness may be difficult to determine. However, instead of using the optimal robustness we instead

use stabilizer psuedomixtures built up from constant size blocks of qubits, here limited to 5 qubits per block. For instance, given $t = bm$ copies of the $|H\rangle$ state, we break it into b blocks of m-qubits $(|H\rangle^{\otimes m})^{\otimes b}$ and so work with a pseudomixture whose sample complexity scales as $\sum_i |x_i| = \mathcal{R}(|H\rangle^{\otimes m})^b = \left[\mathcal{R}(|H\rangle^{\otimes m})^{\frac{1}{m}}\right]^t$.

Numerical results.— We performed substantial numerical investigations up to 5 qubit systems, for which there are over two million stabilizer states and over one thousand Pauli operators. For ascending $m \leq 5$ we calculated $\mathcal{R}(|H^{\otimes m}\rangle)^{\frac{1}{m}}$ as $\{1.414, 1.322, 1.304, 1.301, 1.298\}$. The decrease with m shows a strongly submultiplicative behaviour and reduces simulation overheads (though analytic lower bounds derived in Supplementary Materials show going to higher m cannot reduce this value below 1.207). Specifically, $\mathcal{R}(|H^{\otimes 5}\rangle)^{\frac{2t}{5}} \approx 1.298^{2t} = 1.685^t$ characterises the complexity of our Clifford plus t T gate simulation. This gives exponential improvement over the method used in [19] with complexity scaling as 1.9185^t. A more efficient use of the stabilizer Schmidt decomposition was subsequently established in [18], leading to complexity scaling as 1.385^t, although the restriction to pure states as in [19] also holds here. Even better scaling with t can be obtained by using approximate states [18], but at the price of δ^{-5} overhead in the precision δ. A quantum computation using z CCZ gates, can be implemented using $4z$ T gates [31], implying our simulation complexity scales as 8.067^z. However, as discussed earlier, we can use $|CCZ\rangle$ resource states. We found $\mathcal{R}(|CCZ\rangle) = 2.555$, and so z CCZs can be simulated with an overhead dominated by $\mathcal{R}(|CCZ\rangle)^{2z} = 6.531^z$. This gives exponential improvement over using the four-T-gate gadgetization. Alternative gadgetizations using $|U\rangle = U|+\rangle$ with third-level diagonal U follow naturally and the simulation overhead is given by $\mathcal{R}(|U\rangle)$, see Supplementary Material for many possible examples.

Lower bounds on gate synthesis.— Beyond classical simulation, robustness can help us investigate gate synthesis. Above, we noted that four T gates can exactly synthesize a CCZ gate. How can we be sure that a more complicated or clever scheme does not use even fewer T gates, not just for CCZ gates but more generally? This is an important instance where our resource theory of magic applies; a potential resource state $|U\rangle = U|+\rangle$ for a non-Clifford unitary U cannot be made using t T gates if $\mathcal{R}(|U\rangle) > \mathcal{R}(|H^{\otimes t}\rangle)$. The resource state $|CCZ\rangle$, Clifford-equivalent to a Toffoli resource, has $\mathcal{R}(|CCZ\rangle) = 2.555$ implying

$$\mathcal{R}(|H^{\otimes 3}\rangle) < \mathcal{R}(|CCZ\rangle) < \mathcal{R}(|H^{\otimes 4}\rangle). \qquad (6)$$

Therefore, we establish that the four T gate synthesis of CCZ is optimal. Furthermore, the standard decomposition of CS into 3 T gates (see Fig. 1) is provably optimal since $\mathcal{R}(|H^{\otimes 2}\rangle) < \mathcal{R}(|CS\rangle) < \mathcal{R}(|H^{\otimes 3}\rangle)$. The numerical results in Supplementary Material allow for the inference

of optimality for many more unitaries.

Improved gate synthesis.— Circuit synthesis can be purely unitary over the Clifford+T gate set, or more generally can make use of stabilizer ancillas and measurement thereby using even fewer T gates. The scale of the potential savings is exemplified by the four T gate realisation of CCZ [31], which is assisted by ancillas and measurement. Purely unitary synthesis of CCZ over Clifford+T is known to need at least seven T gates [24]. We shed new light on this phenomena by showing that it emerges from Clifford equivalence of magic states, and give new examples of improved synthesis. The interesting examples arise when $C|U\rangle = |V\rangle$, and yet unitary synthesis of U uses fewer T-gates than V. For these remarkable examples, despite Clifford equivalence of states $|U\rangle$ and $|V\rangle$, there do not exist Cliffords C_1 and C_2 such that $C_1 U C_2 = V$. One explicit example, comparable to Jones' construction [31], starts with the Toff* gate corresponding to $CCZ_{123}CS_{12}$, which is known to be unitarily synthesizable using four T gates [29]. Because the "square-root-of-NOT" $\sqrt{X}$ is a Clifford gate, we also have that

$$\left(II\sqrt{X}\right)|\text{Toff}^*\rangle = |CCZ\rangle. \tag{7}$$

Clifford equivalence of magic states provides an alternative proof that CCZ can be performed with four T gates.

We have found a number of similar examples using the following method (i) Identify U and V with different T-count but whose states have the same robustness $\mathcal{R}(|U\rangle) = \mathcal{R}(|V\rangle)$, (ii) Search for the Clifford C that takes $|U\rangle$ to $|V\rangle$. Existence of such a Clifford is not guaranteed by virtue of $\mathcal{R}(|U\rangle) = \mathcal{R}(|V\rangle)$, but we found a Clifford in every instance investigated. Note also that our T count is over the CNOT+T basis, which is less general than the Clifford+T basis but existing techniques for the latter [24] are impractical for more than three qubits. The two methods (Clifford+T and CNOT+T) give the same T count for CCZ and it is an interesting open question whether they always agree on the T cost of synthesizing a third-level gate. We list, in compact notation, a few of the new synthesis results and the T savings (more are provided in Supplementary Material). For example the CCZ construction discussed above would be represented as

$$CCZ_{123} \xrightarrow{7\to4} CS_{12}CCZ_{123}, \tag{8}$$

where the subscripts denote the qubits on which a third-level gate acts and the numbers above the arrow denote the T cost. Other examples include

$$CCZ_{123} \xrightarrow{7\to4} CS_{12}CS_{13}, \tag{9}$$

$$CCZ_{123,145} \xrightarrow{11\to8} CS_{12,13,14,15}, \tag{10}$$

$$T_{1,2,3}CS_{12,23,13} \xrightarrow{6\to5} T_{2,3}CS_{12,23,13}. \tag{11}$$

Discussion.- By reformulating robustness as an optimization in (2) this facilitates a comparison with recent related works see Table I. For qudit-based computation, Veitch et al. [7] showed that sum-negativity $sn(\rho)$ of a state's discrete Wigner function was a well-defined resource and Pashayan et al. [21] showed how the run time of a Monte-Carlo type sampling algorithm was slower by a factor quadratic in the size of the sum-negativity. In the qudit setting the natural choice for the columns of A in Eq. (2) are the vertices of the Wigner polytope (a larger, but more geometrically simple object than the stabilizer polytope), and phase point operators form a natural operator basis. With these choices, b is a vectorised version of the Wigner representation of ρ and the matrix A becomes the identity matrix. $\mathcal{R}(\rho)$ is simply the sum-negativity (equal to the ℓ_1 norm) of the Wigner quasiprobability distribution associated with ρ. In other words, sum-negativity is just robustness relative to the set of operators with non-negative discrete Wigner function. Unlike our approach, the discrete Wigner-function is not easily adapted to qubits (although see [48, 49]).

In work by Bravyi, Smith and Smolin [19] t-fold copies of $|H\rangle$ are decomposed as linear combinations of stabilizer vectors [20]; the number of terms χ—the stabilizer Schmidt rank—in the decomposition quantifies the simulation overhead. Finding the optimal decomposition is an ℓ_0 minimization ($||x||_0 = |\{i : x_i \neq 0\}|$), which is non-convex and NP-hard, limiting calculations to small number of qubits. Bravyi and Gossett [18] extended this analysis by efficiently finding approximate decompositions that are still sufficient for the task of simulating the outcome of a quantum algorithm. This approximation precludes the possibility of ordering states by the amount of resource, however. We note that it is well known in the signal processing literature [50] that the solution to ℓ_1 minimization also provides a (qualitatively) good solution for ℓ_0 minimization.

Resource	$|\psi\rangle \in \mathbb{C}^d$	$\rho \in B(\mathcal{H})$
$\chi(|\psi\rangle) = ||x||_0$	BSS[19], BG[18]	
$sn(\rho) = ||x||_1$		PWB[21], VMGE[7]
$\mathcal{R}(\rho) = ||x||_1$		This Work

TABLE I. Restatement of related work in terms of norm-minimizing solutions of a system of equations $Ax = b$. The amount of resource in an ancillary state $|\psi\rangle$ or ρ quantifies the classical simulation overhead. In the first and third lines the columns of A are n-qubit stabilizer states (as complex vectors or generalized Bloch vectors, respectively). In the second line the columns of A are extreme points of the Wigner polytope.

In Bravyi, Smith and Smolin [19] it is conjectured that $|H^{\otimes t}\rangle$ has the smallest χ of all t-fold copies of a single-qubit non-stabilizer state. In our work we see that $|H^{\otimes t}\rangle$ have relatively large robustness. This is curious and worthy of further investigation but is also strongly reminiscent of [51] where the (entanglement) Schmidt rank is

seen to disagree with almost every other continuous entanglement measure. A related open problem is to reconcile the fact that small angle ancillae $(1, e^{i\phi\approx0})/\sqrt{2}$ are cheap in our framework, yet are harder to synthesize over the Clifford+T gate [28] set and harder to fault-tolerantly distill [26, 52]. Considerations such as this suggest that a combination of both the stabilizer Schmidt rank and robustness pictures of magic could prove useful.

* m.howard@sheffield.ac.uk
† earltcampbell@gmail.com

[1] F. Brandão and G. Gour, Phys. Rev. Lett. **115**, 070503 (2015).
[2] B. Coecke, T. Fritz, and R. W. Spekkens, Information and Computation (2016), 10.1016/j.ic.2016.02.008.
[3] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, Phys. Rev. Lett. **112**, 120401 (2014).
[4] M. Horodecki and J. Oppenheim, Int. J. Mod. Phys. B **27**, 1345019 (2013).
[5] C. Napoli, T. R. Bromley, M. Cianciaruso, M. Piani, N. Johnston, and G. Adesso, Phys. Rev. Lett. **116**, 150502 (2016).
[6] D. Stahlke, Phys. Rev. A **90**, 022302 (2014).
[7] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, New J. Phys. **16**, 013009 (2014).
[8] G. Vidal and R. Tarrach, Phys. Rev. A **59**, 141 (1999).
[9] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).
[10] E. Knill, Nature **434**, 39 (2005).
[11] E. T. Campbell, B. M. Terhal, and C. Vuillot, arXiv:1612.07330 [quant-ph] (2016), arXiv:1612.07330 [quant-ph].
[12] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature **510**, 351 (2014).
[13] N. Delfosse, C. Okay, J. Bermejo-Vega, D. E. Browne, and R. Raussendorf, arXiv:1610.07093 [quant-ph] (2016), arXiv:1610.07093 [quant-ph].
[14] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, New J. Phys. **14**, 113011 (2012).
[15] A. Mari and J. Eisert, Phys. Rev. Lett. **109**, 230503 (2012).
[16] D. Gross, Journal of Mathematical Physics **47**, 122107 (2006).
[17] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).
[18] S. Bravyi and D. Gosset, Phys. Rev. Lett. **116**, 250501 (2016).
[19] S. Bravyi, G. Smith, and J. A. Smolin, Phys. Rev. X **6**, 021043 (2016).
[20] H. J. García, I. L. Markov, and A. W. Cross, Quantum Information & Computation **14**, 683 (2014).
[21] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Phys. Rev. Lett. **115**, 070501 (2015).
[22] M. Amy, D. Maslov, and M. Mosca, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **33**, 1476 (2014).
[23] M. Amy and M. Mosca, arXiv:1601.07363 [quant-ph] (2016), arXiv:1601.07363 [quant-ph].
[24] D. Gosset, V. Kliuchnikov, M. Mosca, and V. Russo, Quantum Information & Computation **14**, 1261 (2014).
[25] A. Bocharov, Y. Gurevich, and K. M. Svore, Phys. Rev. A **88**, 012313 (2013).
[26] G. Duclos-Cianci and K. M. Svore, Phys. Rev. A **88**, 042325 (2013).
[27] A. Paetznick and K. M. Svore, Quantum Information & Computation **14**, 1277 (2014).
[28] N. J. Ross and P. Selinger, arXiv:1403.2975 [quant-ph] (2014), arXiv:1403.2975 [quant-ph].
[29] P. Selinger, Phys. Rev. A **87**, 042302 (2013).
[30] N. Wiebe and M. Roetteler, Quantum Information and Communication **16**, 134 (2016).
[31] C. Jones, Phys. Rev. A **87**, 022328 (2013).
[32] D. Andersson, I. Bengtsson, K. Blanchfield, and H. B. Dang, J. Phys. A **48**, 345301 (2015).
[33] B. W. Reichardt, Quantum Information & Computation **9**, 1030 (2009).
[34] E. T. Campbell, Phys. Rev. A **83**, 032317 (2011).
[35] A. W. Harrow and M. A. Nielsen, Phys. Rev. A **68**, 012308 (2003).
[36] W. van Dam and M. Howard, Phys. Rev. A **83**, 032310 (2011).
[37] S. G. Hoggar, Geometriae Dedicata **69**, 287 (1998).
[38] M. Grant and S. Boyd, in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, edited by V. Blondel, S. Boyd, and H. Kimura (Springer-Verlag Limited, 2008) pp. 95–110.
[39] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming, Version 2.1* (2014).
[40] This certificate of optimality can be used to obtain a magic witness—an operator whose expectation value with respect to stabilizer states is in the interval $[-1, 1]$ and whose expectation with respect to ρ is $\mathcal{R}(\rho)$. These witnesses can be used to derive exact expressions for robustness as in Supplementary Material.
[41] H. Buhrman, R. Cleve, M. Laurent, N. Linden, A. Schrijver, and F. Unger, in *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)* (2006) pp. 411–419.
[42] S. Virmani, S. F. Huelga, and M. B. Plenio, Physical Review A **71**, 042328 (2005).
[43] M. B. Plenio and S. Virmani, New Journal of Physics **12**, 033012 (2010).
[44] D. Gottesman and I. L. Chuang, Nature **402**, 390 (1999).
[45] X. Zhou, D. W. Leung, and I. L. Chuang, Phys. Rev. A **62**, 052316 (2000).
[46] E. T. Campbell and M. Howard, arXiv:1606.01904 [quant-ph] (2016), arXiv:1606.01904 [quant-ph].
[47] E. T. Campbell and M. Howard, arXiv:1606.01906 [quant-ph] (2016), arXiv:1606.01906 [quant-ph].
[48] N. Delfosse, P. Allard Guerin, J. Bian, and R. Raussendorf, Phys. Rev. X **5**, 021003 (2015).
[49] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega, arXiv:1511.08506 [quant-ph] (2015), arXiv:1511.08506 [quant-ph].
[50] E. Candes and T. Tao, IEEE Transactions on Information Theory **51**, 4203 (2005).
[51] M. Van den Nest, Phys. Rev. Lett. **110**, 060504 (2013).
[52] E. T. Campbell and J. O'Gorman, Quantum Science and Technology **1**, 015007 (2016).

APPENDIX

Robustness of magic

Here we give a more extended discussion of the properties of RoM and also introduce techniques for obtaining lower bounds. We first tackle the basic properties, which are standard for robustness measures of a resource and our proofs will parallel work elsewhere [8]. In contrast, our results on lower bounds (see Sec.) are tailored to the magic state setting and do not have counterparts in other resource theories. Note that robustness is mostly described here in algebraic terms but a geometric picture as in Fig 2 is natural.
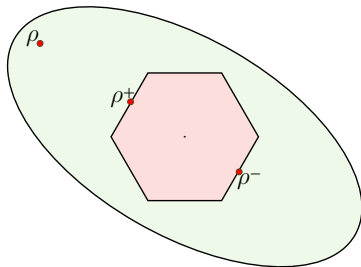

FIG. 2. Robustness of magic (geometric): The hexagon represents all possible mixtures of stabilizer states i.e., the stabilizer polytope. Highlighted points correspond to the state of interest, ρ, and two stabilizer states, $\rho^{\pm}$, such that $\rho = (p+1)\rho^+ - p\rho^-$. Calculating robustness of magic $\mathcal{R}(\rho)$ requires optimizing over all such decompositions but this may be recast as a linear program (Eq. (2) of main text). This is completely analogous to the original formulation of robustness of entanglement [8] except here the extreme points of the set of free states form a discrete set.

Basic properties

We list the basic properties of the robustness of magic

R1 Faithfulness: $\mathcal{R}(\rho) = 1$ if and only if ρ is a stabilizer state, and otherwise $\mathcal{R}(\rho) > 1$.

R2 Submultiplicativity: $\mathcal{R}(\rho_1 \otimes \rho_2) \leq \mathcal{R}(\rho_1) \cdot \mathcal{R}(\rho_2)$;

R3 Monotonicity: for all trace-preserving stabilizer channels $\mathcal{E}$, we have $\mathcal{R}(\mathcal{E}(\rho)) \leq \mathcal{R}(\rho)$.

R4 Convexity: $\mathcal{R}(\sum_k p_k \rho_k) \leq \sum_k |p_k| \mathcal{R}(\rho_k)$.

Let us prove these properties one by one.
Faithfulness.- If ρ is a mixed-stabilizer state then by definition there exists a decomposition $\rho = \sum_i p_i \sigma_i$ where p_i are positive and so setting $x_i = p_i$ we have $\sum_i |x_i| = \sum_i x_i = 1$. Conversely, if ρ is nonstabilizer state, then there is at least one negative value of x_i. Furthermore, $\text{Tr}(\rho) = 1$ entails $\sum_i x_i = 1$ and so we can express the

absolute sum as $\sum_i |x_i| = 1 + 2\sum_{i:x_i<0} |x_i|$ where the new summation is only over the negative values. We see that if a single value is negative, then the RoM exceeds unity.

Submutliplicativity.- Let ρ_1 and ρ_2 have optimal pseudomixtures

$$\rho_1 = \sum_i x_i \sigma_i, \tag{12}$$

$$\rho_2 = \sum_j y_j \sigma_j,$$

where optimality entails $\mathcal{R}(\rho_1) = \sum_i |x_i|$ and $\mathcal{R}(\rho_2) = \sum_j |y_j|$. We consider the composite state

$$\rho_1 \otimes \rho_2 = \sum_{i,j} x_i y_j (\sigma_i \otimes \sigma_j). \tag{13}$$

Since $\sigma_i \otimes \sigma_j$ are pure stabilizer states, this is a stabilizer pseudomixture with quasiprobabilities $x_i y_j$. Taking the absolute sum we have

$$\sum_{i,j} |x_i y_j| = \left(\sum_i |x_i|\right) \left(\sum_j |y_j|\right) \tag{14}$$

$$= \mathcal{R}(\rho_1)\mathcal{R}(\rho_2).$$

Therefore, $\mathcal{R}(\rho_1 \otimes \rho_2) \leq \mathcal{R}(\rho_1)\mathcal{R}(\rho_2)$.
Monotonicity.- If $\mathcal{E}$ is stabilizer operation, then $\mathcal{E}(\sigma_i) = \sum_j p_{i,j}\sigma_j$ where $p_{i,j} \geq 0$. Furthermore, if $\mathcal{E}$ is trace-preserving then, for all i, we have $\sum_j p_{i,j} = 1$. Combined with positivity we know $\sum_j |p_{i,j}| = 1$. Therefore, applying $\mathcal{E}$ to an arbitrary state

$$\mathcal{E}(\rho) = \sum_i x_i \sum_j p_{i,j}\sigma_j, \tag{15}$$

$$= \sum_j \left(\sum_i x_i p_{i,j}\right)\sigma_j,$$

which is a stabilizer pseudomixture with quasiprobabilities $x'_j = \sum_i x_i p_{i,j}$. Therefore,

$$\mathcal{R}(\mathcal{E}(\rho)) \leq \sum_j |\sum_i x_i p_{i,j}|. \tag{16}$$

Using the triangle inequality $|\sum_k a_k| \leq \sum_k |a_k|$ and $|ab| = |a| \cdot |b|$, we have

$$\mathcal{R}(\mathcal{E}(\rho)) \leq \sum_i |x_i| \sum_j |p_{i,j}| \tag{17}$$

$$= \sum_i |x_i|$$

$$= \mathcal{R}(\rho).$$

By similar methods we can show that the average robustness is non-increasing under a trace non-increasing map

e.g. postselection on a particular outcome of a stabilizer POVM. The proof proceeds along the same lines as that provided in [8].

Convexity.- Consider a set of quantum states with optimal stabilizer pseudomixtures

$$\rho_k = \sum_i x_i^{(k)} \sigma_i, \tag{18}$$

so that $\mathcal{R}(\rho_k) = \sum_i |x_i^{(k)}|$. It follows that

$$\sum_k p_k \rho_k = \sum_i \sum_k p_k x_i^{(k)} \sigma_i. \tag{19}$$

Taking $x_i' = \sum_k p_k x_i^{(k)}$ as quasiprobabilities we have

$$\mathcal{R}\left(\sum_k p_k \rho_k\right) \le \sum_i |\sum_k p_k x_i^{(k)}|. \tag{20}$$

Applying the triangle inequality, we have

$$\mathcal{R}\left(\sum_k p_k \rho_k\right) \le \sum_i \sum_k |p_k| \cdot |x_i^{(k)}| \tag{21}$$
$$= \sum_k |p_k| \mathcal{R}(\rho_k).$$

This proves convexity.

Taking the logarithm of the robustness creates a related measure LR such that $LR(\rho) = \log_2(\mathcal{R}(\rho))$. It immediately follows that LR is also monotonically decreasing under stabilizer channels. Submultiplicativity translates into subadditivity. Furthermore, LR is faithful in the sense that $LR(\rho) = 0$ if and only if ρ is a stabilizer state.

Lower bounds on RoM

This section introduces a new quantity that we use to establish lower bounds on RoM. Such bounds are valuable since numerical methods are limited to modest numbers of qubits, whereas the lower bounds will hold for any number of qubits.

We define $\mathcal{D}$ (also called the st-norm $||\cdots||_{\mathrm{st}}$ in Ref. [34]) as

$$\mathcal{D}(\rho) = \frac{1}{2^n} \sum_{P \in \mathcal{P}_+} |\mathrm{Tr}(P\rho)|, \tag{22}$$

where $\mathcal{P}_+$ is the set of Pauli operators with $+1$ phase, including the identity. It has some useful properties,

D1 Convexity: $\mathcal{D}(\sum_k p_k \rho_k) \le \sum_k |p_k| \mathcal{D}(\rho_k)$,

D2 Magic witness: if $\mathcal{D}(\rho) > 1$ then ρ is a nonstabilizer state,

D3 Lower bound: $\mathcal{D}(\rho) \le \mathcal{R}(\rho)$,

D3* Tighter lower bound: If ρ is an n-qubit state then

$$\frac{\mathcal{D}(\rho) - \frac{1}{2^n}}{\left(1 - \frac{1}{2^n}\right)} \le \mathcal{R}(\rho), \tag{23}$$

D4 Multiplicativity: $\mathcal{D}(\rho_1 \otimes \rho_2) = \mathcal{D}(\rho_1) \cdot \mathcal{D}(\rho_2)$.

Convexity.- We have by linearity of the trace that

$$\mathcal{D}(\rho) = \frac{1}{2^n} \sum_{P \in \mathcal{P}_+} |\mathrm{Tr}[P \sum_k p_k \rho_k]| \tag{24}$$
$$= \frac{1}{2^n} \sum_{P \in \mathcal{P}_+} |\sum_k \mathrm{Tr}[P p_k \rho_k]|.$$

The triangle inequality $|\sum_k a_k| \le \sum_k |a_k|$ entails

$$\mathcal{D}(\rho) \le \frac{1}{2^n} \sum_{P \in \mathcal{P}_+} \sum_k |\mathrm{Tr}[P p_k \rho_k]|. \tag{25}$$

A factor $|p_k|$ can come outside the absolute value sign, and reordering the summations we have

$$\mathcal{D}(\rho) \le \frac{1}{2^n} \sum_k |p_k| \sum_{P \in \mathcal{P}_+} |\mathrm{Tr}[P \rho_k]| \tag{26}$$
$$= \sum_k |p_k| \mathcal{D}(\rho_k),$$

which proves convexity.

Magic witness.- A pure n-qubit stabilizer state σ_i has a stabilizer group $G_{\sigma_i} \subset \mathcal{P}$ (the Pauli group on n qubits) containing 2^n elements, so that

$$\sigma_i = \frac{1}{2^n} \sum_{g \in G_{\sigma_i}} g. \tag{27}$$

and such a pure state must have $\mathcal{D}(\sigma_i) = 1$. For a mixed stabilizer state, $\rho = \sum_i p_i \sigma_i$ with $p_i > 0$, and we can apply convexity so that

$$\mathcal{D}(\rho) \le \sum_i p_i \mathcal{D}(\sigma_i) \tag{28}$$
$$= \sum_i p_i = 1.$$

Since all stabilizer states have $\mathcal{D}(\rho) \le 1$, this implies that if $\mathcal{D}(\rho) > 1$ then ρ must be a nonstabilizer state. Note that some mixed stabilizer states will have $\mathcal{D}(\rho) < 1$.

Lower bound.- Let $\rho = \sum_i x_i \sigma_i$ be an optimal stabilizer pseudomixture. By convexity we have

$$\mathcal{D}(\rho) \le \sum_i |x_i| \mathcal{D}(\sigma_i), \tag{29}$$

and using $\mathcal{D}(\sigma_i) = 1$ we have

$$\mathcal{D}(\rho) \le \sum_i |x_i| = \mathcal{R}(\rho), \tag{30}$$

which completes the proof.

Tighter lower bound.- Next, we prove an even tighter lower bound. Asymptotically, the tighter bound is identical to the above lower bound, but it is much tighter for modest numbers of qubits. We begin by revisiting the first line of the convexity proof and split the sum over $\mathcal{P}_+$ into one term for $\mathbb{1} \in \mathcal{P}_+$ and the remainder $\mathcal{P}^* = \mathcal{P}_+ \setminus \mathbb{1}$ so that

$$\mathcal{D}(\rho) = \frac{1}{2^n}\left(|\mathrm{Tr}[\rho]| + \sum_{P \in \mathcal{P}^*} |\mathrm{Tr}[P\rho]| \right) \qquad (31)$$

$$= \frac{1}{2^n}\left(1 + \sum_{P \in \mathcal{P}^*} |\mathrm{Tr}[P\rho]| \right)$$

Writing ρ as its optimal stabilizer pseudomixture and applying the triangle inequality we arrive at

$$\mathcal{D}(\rho) \leq \frac{1}{2^n}\left(1 + \sum_i |x_i| \sum_{P \in \mathcal{P}^*} |\mathrm{Tr}[P\sigma_i]| \right). \qquad (32)$$

We note that

$$\frac{1}{2^n}\sum_{P \in \mathcal{P}^*} |\mathrm{Tr}[P\sigma_i]| = \frac{1}{2^n}\left(\sum_{P \in \mathcal{P}_+} |\mathrm{Tr}[P\sigma_i]| \right) - \frac{1}{2^n} \qquad (33)$$

$$= \mathcal{D}(\sigma_i) - \frac{1}{2^n}$$

$$= 1 - \frac{1}{2^n}.$$

Applying this to Eq. (31) gives

$$\mathcal{D}(\rho) \leq \frac{1}{2^n} + \left(\sum_i |x_i| \right)\left(1 - \frac{1}{2^n} \right). \qquad (34)$$

Using $\mathcal{R}(\rho) = \sum_i |x_i|$ and rearranging for $\mathcal{R}(\rho)$, we find the tighter lower bound stated earlier.

Multiplicativity.- For a product state we have

$$\mathcal{D}(\rho_1 \otimes \rho_2) = \frac{1}{2^n}\sum_{\mathcal{P} \in \mathcal{P}_+} |\mathrm{Tr}[P(\rho_1 \otimes \rho_2)]|. \qquad (35)$$

Every $P \in \mathcal{P}_+$ can be written as a product $P = P_1 \otimes P_2$ where $P_1 \in \mathcal{P}_+$ and $P_2 \in \mathcal{P}_+$ for the smaller Hilbert spaces. Therefore

$$\mathcal{D}(\rho_1 \otimes \rho_2) = \frac{1}{2^n}\sum_{P_1, P_2 \in \mathcal{P}_+} |\mathrm{Tr}[(P_1\rho_1) \otimes (P_2\rho_2)]|. \qquad (36)$$

The trace is multiplicative with respect to tensor products ($\mathrm{Tr}[A \otimes B] = \mathrm{Tr}[A]\mathrm{Tr}[B]$) and so

$$\mathcal{D}(\rho_1 \otimes \rho_2) = \frac{1}{2^n}\sum_{P_1, P_2 \in \mathcal{P}_+} |\mathrm{Tr}[(P_1\rho_1)| \cdot |\mathrm{Tr}[(P_2\rho_2)]| \qquad (37)$$

$$= \sum_{P_1, P_2 \in \mathcal{P}_+} \frac{1}{2^{n_1}}|\mathrm{Tr}[(P_1\rho_1)| \cdot \frac{1}{2^{n_2}}|\mathrm{Tr}[(P_2\rho_2)]|$$

$$= \mathcal{D}(\rho_1)\mathcal{D}(\rho_2),$$

where in the second line we split $n = n_1 + n_2$ where n_1 and n_2 are the number of qubits in systems 1 and 2. This proves multiplicativity.

The combination of multiplicativity with the lower bound property entails that $\mathcal{D}(\rho)^n = \mathcal{D}(\rho^{\otimes n}) \leq \mathcal{R}(\rho^{\otimes n})$. This enables us to lower bound robustness for large n even though numerically finding the robustness is difficult for large n. Let us apply these techniques to mixed versions of the magic states singled out by Bravyi and Kitaev [9] (we replace their T with F to avoid notational confusion),

$$\rho_H(r) = \frac{1}{2}\left(\mathbb{1} + r\frac{X + Z}{\sqrt{2}} \right), \qquad (38)$$

$$\rho_F(r) = \frac{1}{2}\left(\mathbb{1} + r\frac{X + Y + Z}{\sqrt{3}} \right), \qquad (39)$$

where $r = \pm 1$ for pure states and $-1 < r < +1$ for mixed states. It is easy to verify that

$$\mathcal{D}(\rho_H(r)) = \frac{1}{2}(1 + \sqrt{2}r), \qquad (40)$$

$$\mathcal{D}(\rho_F(r)) = \frac{1}{2}(1 + \sqrt{3}r). \qquad (41)$$

Therefore, for pure states $r = 1$, we can conclude

$$1.207^n \leq \mathcal{R}(\rho_H(1)^{\otimes n}), \qquad (42)$$

$$1.366^n \leq \mathcal{R}(\rho_F(1)^{\otimes n}). \qquad (43)$$

Using multiplicativity with the tighter lower bound, for a m-qubit state ρ we have

$$\frac{\mathcal{D}(\rho)^n - \frac{1}{2^{nm}}}{\left(1 - \frac{1}{2^{nm}} \right)} \leq \mathcal{R}(\rho^{\otimes n}). \qquad (44)$$

Whenever $\mathcal{D}(\rho) > 1$, the lower bound clearly approaches $\mathcal{D}(\rho)^n$ as $n \to \infty$.

Robustness of Particular States

Robustness of States from the Third Level of the Clifford Hierarchy

In the main text we gave the numerical value for the robustness of t-fold copies of $|H\rangle$. Here we provide the exact symbolic expressions for small t,

$$\mathcal{R}(|H\rangle) = \sqrt{2} \approx 1.4142, \qquad (45)$$

$$\mathcal{R}(|H^{\otimes 2}\rangle) = \frac{1 + 3\sqrt{2}}{3} \approx 1.7476, \qquad (46)$$

$$\mathcal{R}(|H^{\otimes 3}\rangle) = \frac{1 + 4\sqrt{2}}{3} \approx 2.2190, \qquad (47)$$

$$\mathcal{R}(|H^{\otimes 4}\rangle) = \frac{3 + 8\sqrt{2}}{5} \approx 2.8627. \qquad (48)$$

We also have

$$\mathcal{R}(|H^{\otimes 5}\rangle) \approx 3.68705 \qquad (49)$$

although without a neat symbolic expression.

For t up to 5 we have calculated expressions for the exact value of $\mathcal{R}(|H^{\otimes t}\rangle)$. For $t > 5$ we can also place bounds on this quantity as shown in Table II. In addition, we can give a full classification of all 3-qubit diagonal gates from the set generated by CNOT+T as in Table III. A similar table can be constructed for 4-qubit diagonal gates using the data provided in Supplementary Material.

Numerical maximization of Robustness

Given that we have an operationally relevant quantifier of magic, it is interesting to consider for which states ρ the quantity $\mathcal{R}(\rho)$ is maximized. For a single qubit the state $|F\rangle$ [9] with Bloch vector $(x, y, z) = (1, 1, 1)/\sqrt{3}$ given in (39) has maximal robustness $\sqrt{3}$. Multiple copies of this state have robustness

$$\mathcal{R}(|F^{\otimes 2}\rangle) = \frac{1 + 2\sqrt{3}}{2} \approx 2.232, \qquad (50)$$

$$\mathcal{R}(|F^{\otimes 3}\rangle) = \frac{1 + 3\sqrt{3}}{2} \approx 3.098, \qquad (51)$$

$$\mathcal{R}(|F^{\otimes 4}\rangle) = \frac{13 + 20\sqrt{3}}{11} \approx 4.331. \qquad (52)$$

For two qubits, robustness is maximized at $\sqrt{5}$ by a state that is maximally outside one of the facets of the 2-qubit stabilizer polytope (Table 2 Line 5 of [33]). The state $(1, 1, 1, i)/2$ is maximally robust amongst all "flat" (equally-weighted) 2-qubit states achieving $\mathcal{R} = 2.2$. For 3 qubits the most robust state at $\mathcal{R} = 3.8$ is the so-called Hoggar state [37],

$$|\text{Hoggar}\rangle = \frac{1}{\sqrt{6}} \begin{pmatrix} 1+i \\ 0 \\ -1 \\ 1 \\ -i \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad (53)$$

a fiducial vector for a symmetric, informationally complete positive operator valued measure (SIC-POVM) co-variant under the 3-qubit Pauli group. Some of these states were already classified as highly non-stabilizer by other means [32]. In Fig. 3 we depict the robustness of a one-parameter family of m-qubit states, for $m \in \{1, 2, 3\}$. From this figure we can read off $\mathcal{R}(|H^{\otimes 2}\rangle) < \mathcal{R}(|CS\rangle) < \mathcal{R}(|H^{\otimes 3}\rangle)$ and $\mathcal{R}(|H^{\otimes 3}\rangle) < \mathcal{R}(|CCZ\rangle) < \mathcal{R}(|H^{\otimes 4}\rangle)$. We can also see that a doubly-controlled rotation of angle

$\theta = \pi/2$—a CCS gate—would require at least 5 T gates to implement.

By exploiting the Jamiołkowski isomorphism, as in e.g. [35, 36], we can also investigate the robustness of operations. In particular, for every unitary $U \in \mathrm{U}(2^n)$ we can associate a state $|J_U\rangle = (\mathbb{1} \otimes U) \sum_{j \in \mathbb{Z}_2^n} |j, j\rangle$. Maximizing the quantity $\mathcal{R}(|J_U\rangle)$ tells us the most robust unitary operations. For a single qubit we numerically find the optimal unitary to be

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & e^{i\frac{3\pi}{4}} \\ e^{-i\frac{\pi}{4}} & 1 \end{pmatrix}, \qquad (54)$$

$$\mathcal{R}(|J_U\rangle) = \mathcal{R}(|H^{\otimes 2}\rangle) = \frac{1 + 3\sqrt{2}}{3}, \qquad (55)$$

whereas optimizing over two-qubit unitaries we find

$$U = \frac{1}{5} \begin{pmatrix} -1 - 2i & 3 + i & 1 - 3i & 0 \\ 1 - 3i & -3 - i & -1 - 2i & 0 \\ 3 + i & 1 + 2i & -3 - i & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}, \qquad (56)$$

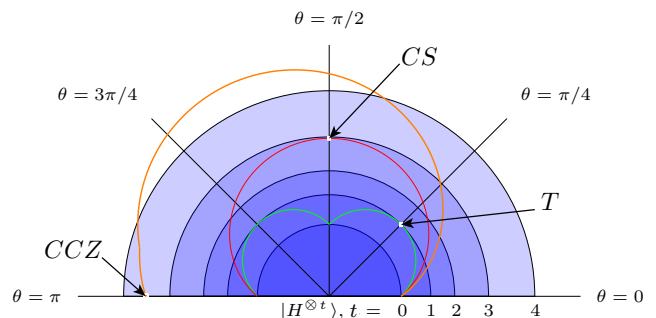$$\mathcal{R}(|J_U\rangle) = \frac{23}{5}. \qquad (57)$$

FIG. 3. Polar plot where the radial coordinate denotes robustness and the angular component represents the last entry in a m-qubit resource state $(1, \ldots, e^{i\theta})^T / 2^{m/2}$. Concentric semi-circles correspond to robustness $\mathcal{R}(|H^{\otimes t}\rangle)$ for $t = 0, 1, \ldots, 4$. Highlighted points correspond to resource states for T, CS and CCZ. We see that $\mathcal{R}(|H^{\otimes 3}\rangle)$ is only slightly greater than $\mathcal{R}(|CS\rangle)$, meaning that CS synthesis using 3 T gates as in Fig. 1(c) of main text is optimal and no ancilla-assisted or non-deterministic strategy could be more parsimonious.

Interconvertability

Here we present some further examples of T-saving synthesis of diagonal gates from the CNOT+T gate set. The savings come from Clifford equivalence of magic states as discussed in the main text in the context of

	Lower from $\mathcal{D}$ (23) and (40)	Lower from $\mathcal{R}(|U\rangle)$	$U \in \langle \text{CNOT}, T\rangle$	Upper
$\mathcal{R}(|H^{\otimes 6}\rangle)$	3.1269	4.7031	$T_1 CS_{24,35,45}$	4.9238
$\mathcal{R}(|H^{\otimes 7}\rangle)$	3.75592	5.5242	$T_4 CS_{13,24,35,45}$	6.3523
$\mathcal{R}(|H^{\otimes 8}\rangle)$	4.52157	5.7934	$T_5 CS_{12,35,45}$	8.1953
$\mathcal{R}(|H^{\otimes 9}\rangle)$	5.4501	6.2625	$T_{1,2,3} CS_{12,14,25} CCZ_{345}$	10.555
$\mathcal{R}(|H^{\otimes 10}\rangle)$	6.5738	6.8995	$T_1 CS_{12,23,14,25,45} CCZ_{345}$	13.594
$\mathcal{R}(|H^{\otimes 11}\rangle)$	7.9321	6.9255	$T_5 CS_{23,24,45} CCZ_{125,345}$	17.341

TABLE II. Bounds on $\mathcal{R}(|H^{\otimes t}\rangle)$ for $6 \le t \le 11$: The upper bound is given by $\mathcal{R}(|H^{\otimes t}\rangle) \le \mathcal{R}(|H^{\otimes \lfloor \frac{t}{2}\rfloor}\rangle)\mathcal{R}(|H^{\otimes \lceil \frac{t}{2}\rceil}\rangle)$ which holds because of submultiplicativity of robustness. The first lower bound derives from expression (23) using $\mathcal{D}(|H^{\otimes t}\rangle) = [\mathcal{D}(|H\rangle)]^t = \left[\frac{1+\sqrt{2}}{2}\right]^t$ (40). The second lower bound corresponds to $\mathcal{R}(|U\rangle)$ where U, provided in the column headed $U \in \langle \text{CNOT}, T\rangle$, is synthesizable over the CNOT+T gate set using t T gates [23]. Clearly any such U satisfies $\mathcal{R}(|U\rangle) \le \mathcal{R}(|H^{\otimes t}\rangle)$. Another way of interpreting this column is to note that these U are the most robust gates we know how to construct over the CNOT+T gate set using $t \in \{6, 7, \ldots, 11\}$ T gates. There are no five-qubit $U \in \langle \text{CNOT}, T\rangle$ that require more than 11 T gates.

$\mathcal{R}(|U\rangle)$	T cost	U	T cost$'$	U'
1.41421	1	T_1		
1.74755	2	$T_{1,2}$		
2.2	3	CS_{12}		
2.21895	3	$T_{1,2,3}$		
2.55556	4	$CS_{12,13}$	7	CCZ_{123}
2.80061	4	$T_1 CS_{23}$		
3.12132	5	$T_1 CS_{12,13}$	6	$T_1 CCZ_{123}$

TABLE III. Three qubit classification: All the three-qubit diagonal gates from the third level of the Clifford hierarchy can be classified by the robustness of the associated resource state $|U\rangle = U|+\rangle$. The cost of synthesizing these gates over the Clifford+T gate set is provided. Two gates U and U' with the same robustness and different T cost are equivalent via the construction described in connection with Eq. (7) of the main text.

equation (7);

$$T_{1,2} CCZ_{345} \xrightarrow{8\to 6} T_{1,2} CS_{35,45}, \quad (58)$$

$$T_{1,2,5} CCZ_{345} \xrightarrow{8\to 7} T_{1,2,5} CS_{35,45}, \quad (59)$$

$$T_{1,2,3,4} CS_{23,24,34} \xrightarrow{7\to 6} T_{1,4} CS_{24,34}, \quad (60)$$

$$CS_{12} CCZ_{345} \xrightarrow{9\to 7} CS_{12,35,45}, \quad (61)$$

$$T_5 CS_{12,25} CCZ_{345} \xrightarrow{9\to 7} T_5 CS_{14,25,35,45}. \quad (62)$$

These are all provably optimal in the sense that the T cost on the right hand side is the minimum possible, as it is the smallest integer t satisfying $\mathcal{R}(|U\rangle) \le \mathcal{R}(|H^{\otimes t}\rangle)$.

We also make a brief comment on resource interconvertability. Are there optimal transformations that convert n-qubit $|\phi\rangle$ to m-qubit $|\psi\rangle$, $m \ne n$, with $\mathcal{R}(|\phi\rangle) = \mathcal{R}(|\psi\rangle)$? At least one such transformation does exist. In the range $0 \le \theta \le \arctan \frac{1}{3}$, two copies of the equatorial state $|\phi\rangle = (|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$ have robustness

$$\mathcal{R}(|\phi^{\otimes 2}\rangle) = (2\sin\phi + \sin 2\phi + \cos 2\phi + 1)/2$$

but a measurement of Pauli operator ZY on $|\phi^{\otimes 2}\rangle$, followed by a CNOT, creates the state $|\psi\pm\rangle|Y\pm\rangle$ depending on the ZY outcome, where $|\psi\pm\rangle$ has Bloch vector $\vec{r} = (\cos^2(\theta), \sin(\theta)\cos(\theta), \pm\sin(\theta))$ hence $||\vec{r}||_1 = \mathcal{R}(|\psi\pm\rangle) = \mathcal{R}(|\phi^{\otimes 2}\rangle)$. Finding further examples of such transformations could find application in magic state preparation. The foregoing leads naturally to the question of asymptotic interconvertibility, much studied in the context of entanglement, where one seeks to maximize the rate m/n in the limit $n \to \infty$ of a stabilizer operation mapping $|\phi^{\otimes n}\rangle \to |\psi^{\otimes m}\rangle$. In this work we have dealt solely with deterministic gate synthesis but the extension to probabilistic (depending on measurement outcomes) circuit synthesis is natural; the expected robustness $\mathbb{E}[\mathcal{R}(|\phi\rangle)]$ is now the relevant quantity to compare to e.g, $\mathcal{R}(|H^{\otimes n}\rangle)$. Our choice to focus on synthesizing gates from the third level of the Clifford hierarchy was motivated by the fact that any measurement outcomes m_i in Fig. (1) of main text can be corrected for with a Clifford gate. Other resource states will not have this property and unwanted measurement outcomes could result in either disposing of the output state or correcting with a non-Clifford gate whose cost would also need to be taken into account.