Quantifying magic for multi-qubit operations

James R. Seddon[1] and Earl T. Campbell[2]

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

[2]*Department of Physics and Astronomy, University of Sheffield, Sheffield, UK*
(Dated: 2nd July, 2019)

The development of a framework for quantifying "non-stabiliserness" of quantum operations is motivated by the magic state model of fault-tolerant quantum computation, and by the need to estimate classical simulation cost for noisy intermediate-scale quantum (NISQ) devices. The robustness of magic was recently proposed as a well-behaved magic monotone for multi-qubit states and quantifies the simulation overhead of circuits composed of Clifford+T gates, or circuits using other gates from the Clifford hierarchy. Here we present a general theory of the "non-stabiliserness" of quantum operations rather than states, which are useful for classical simulation of more general circuits. We introduce two magic monotones, called channel robustness and magic capacity, which are well-defined for general n-qubit channels and treat all stabiliser-preserving CPTP maps as free operations. We present two complementary Monte Carlo-type classical simulation algorithms with sample complexity given by these quantities and provide examples of channels where the complexity of our algorithms is exponentially better than previous known simulators. We present additional techniques that ease the difficulty of calculating our monotones for special classes of channels.

1. INTRODUCTION

The Gottesman-Knill theorem showed that circuits comprised of stabiliser state preparations, Clifford gates, Pauli measurements, classical randomness and conditioning can be efficiently simulated by a traditional computer [1, 2]. If a circuit involves a relatively small proportion of non-Clifford operations, simulation may be within the reach of a classical computer, albeit with a runtime overhead that is expected to scale exponentially with the amount of resource required.

An important class of devices comprises so-called near-Clifford circuits where simulation may be feasible [3, 4]. There are two scenarios where near-Clifford circuits are relevant. As we enter the era of Noisy Intermediate Scale Quantum (NISQ) devices [5], many experiments proposed as demonstrators of quantum advantage may be near-Clifford so it is important to rigorously understand when a classical simulation is available. Furthermore, in the NISQ regime the need for classical simulation tools for benchmarking and verification becomes more pressing. The quantification of non-stabiliser resource is also of interest in the context of the magic state model of fault-tolerant quantum computation [6–8], the second scenario. Any device intended to provide quantum advantage must involve non-stabiliser operations. In circuits employing error-correcting codes, however, it is often not possible for the code to 'natively' implement non-Clifford gates fault-tolerantly [8]. Instead, these gates are implemented indirectly by injection of so-called magic states. These are non-stabiliser states that must be prepared using the experimentally costly process of magic state distillation [6–17], which is comprised of Clifford-dominated circuits.

Both of these scenarios motivate the development of a resource theory [17–26] where the class of free operations is generated by stabiliser state preparations and rounds of stabiliser operations as described above. For the case of odd d-dimensional qudits this problem is largely solved by the discrete phase space formalism [27–32]; odd dimension qudit stabiliser states are characterised by a positive discrete Wigner function. In Ref. [33], the discrete Wigner function was cast as

---
* james.seddon.15@ucl.ac.uk

a quasiprobability distribution, making a direct connection between the negativity of the distribution, and the complexity of calculating expectation values via a Monte Carlo-type simulation algorithm. These techniques have recently been extended to quantify the magic of odd-dimension qudit quantum channels [34]. However, the discrete phase space approach cannot be applied cleanly to qubits without excluding some Clifford operations from the free operations [35, 36], or losing the ability to compose representations under tensor product [37]. To retain all multi-qubit stabiliser channels as free operations, then, we must seek alternative approaches.

Howard and Campbell [38] introduced a scheme where density matrices are decomposed as real linear combinations of pure stabiliser state projectors. Non-stabiliser states sit outside the convex hull of the pure stabiliser states, so their decompositions necessarily contain negative terms and can again be viewed as quasiprobability distributions, with ℓ_1-norm strictly larger than 1. The robustness of magic for a state, defined as the minimum ℓ_1-norm over all valid decompositions, is a monotone under stabiliser operations and has several useful resource-theoretic properties. Alternative approaches include stabiliser rank methods, where the state vector is decomposed as a superposition of stabiliser states [39–42]. Exact and approximate stabiliser rank, and the associated quantity extent, are measures of magic for pure states. Here we are interested in measures naturally suited for applications to mixed states or general, noisy quantum channels. A stabiliser-based method to simulate noisy circuits by decomposition of states into Pauli operators was recently proposed in Ref. [43]. In this work we characterise the cost of quantum *operations* with respect to the resource theory of magic. Robustness of magic naturally quantifies the cost for a subclass of non-Clifford operations, namely gates from the third level of the Clifford hierarchy. It is less clear how the framework can be extended to more general quantum operations, and formalising this is one of our main aims.

In Ref. [3], Bennink et al. presented an algorithm in which completely positive trace-preserving (CPTP) maps are decomposed as quasiprobability distributions over a subset of stabiliser-preserving operations that we will call CPR. This subset supplements the Clifford unitaries with Pauli reset channels, in which measurement of some Pauli observable is followed by a conditional Clifford correction, so as to reset a state to a particular +1 Pauli eigenstate. While Bennink et al. showed that CPR spans the set of CPTP maps, there is no guarantee that all stabiliser-preserving CPTP maps can be found within its convex hull. Indeed, we will see in Section 7 there exist channels that are stabiliser-preserving, but are nevertheless assigned a non-trivial cost by the algorithm of Ref. [3]. The implication is that decomposition in terms of elements of CPR is not the best strategy for simulating general non-stabiliser operations. An obvious extension of Ref. [3] is to replace CPR by the full set of stabiliser-preserving CPTP maps. The technical question to be answered is then how to correctly and concisely represent this set; how can we be sure that we have captured all possible stabiliser-preserving channels? This issue is addressed in Sections 4 and 5.

In this paper we introduce two magic monotones for channels: the channel robustness $\mathcal{R}_*$ and the magic capacity $\mathcal{C}$. Both are closely related to the robustness of magic for states. They are well-defined for general n-qubit channels and treat all stabiliser-preserving CPTP maps as free operations. We will see that these monotones give the sample complexity of two classical simulation algorithms. Other magic monotones have been proposed [17, 30, 31] but without known connections to classical simulation algorithms. Furthermore, we give several examples of channels where the simulation complexities of our approaches are exponentially faster (as a function of gate count) than other quasiprobability simulators such as the Bennink et al. simulator [3]. To our knowledge, our algorithms are the first that are known to efficiently simulate all stabiliser-preserving CPTP maps, as opposed to the set of stabiliser operations generated by Clifford gates and Pauli measurements.

The paper is structured as follows. In Section 2 we review the properties of robustness of magic and give some definitions. Next, we summarise our main results in Section 3, before pinning down what we mean by stabiliser-preserving operations in Section 4. Sections 5 and 6 are chiefly con-

cerned with proving important properties of our monotones. Two classical simulation algorithms, each related to one of our monotones, are described in Section 7. Finally, in Section 8 we calculate the numerical values of our monotones for operations on up to five qubits, using techniques developed in Appendix E. MATLAB code to calculate each of our measures is provided at the public repository Ref. [44].

2. PRELIMINARIES

Let STAB_n be the set of n-qubit stabiliser states. In an abuse of notation we will use $|\phi\rangle \in \mathrm{STAB}_n$ to mean a pure state from this set, and $\rho \in \mathrm{STAB}_n$ to mean the density matrix of a state taken from the stabiliser polytope, the convex hull of pure stabiliser states. The pure states in STAB_n form an overcomplete basis for the set of 2^n-dimensional density matrices $\mathcal{D}_n$. We can therefore write the density matrix for *any* state as an affine combination of pure stabiliser state projectors $\rho = \sum_j q_j |\phi_j\rangle\langle\phi_j|$ where $|\phi_j\rangle \in \mathrm{STAB}_n$, and $\sum_j q_j = 1$. In general, q_j can be negative. The robustness of magic is defined as the minimal ℓ_1-norm $\|\vec{q}\|_1 = \sum_j |q_j|$ over all possible decompositions [38]:

$$\mathcal{R}(\rho) = \min_{\vec{q}} \left\{ \|\vec{q}\|_1 : \sum_j q_j |\phi_j\rangle\langle\phi_j| = \rho, \ |\phi_j\rangle \in \mathrm{STAB}_n \right\}. \tag{1}$$

In the definition above, the state of interest is expressed as a decomposition over pure stabiliser states. By collecting together all terms of the same sign, any state can instead be expressed in terms of a pair of mixed stabiliser states (Figure 1). An equivalent definition is then:

$$\mathcal{R}(\rho) = \min_{\rho_\pm \in \mathrm{STAB}_n} \{1 + 2p : (1 + p)\rho_+ - p\rho_- = \rho, p \geq 0\}. \tag{2}$$

The robustness of magic is a well-behaved magic monotone, having the following properties:

1. *Convexity*: $\mathcal{R}\left(\sum_j q_j\rho_j\right) \leq \sum_j |q_j|\mathcal{R}(\rho_j)$;

2. *Faithfulness*: If $\rho \in \mathrm{STAB}_n$, then $\mathcal{R}(\rho) = 1$. Otherwise $\mathcal{R}(\rho) > 1$;

3. *Monotonicity under stabiliser operations*: If Λ is a CPTP stabiliser-preserving operation, then $\mathcal{R}(\Lambda(\rho)) \leq \mathcal{R}(\rho)$;

4. *Submultiplicativity under tensor product*: $\mathcal{R}(\rho^A \otimes \rho'^B) \leq \mathcal{R}(\rho^A)\mathcal{R}(\rho'^B)$.

The quantity $\mathcal{R}$ also has a clear operational meaning, quantifying the classical simulation cost in a Monte Carlo-type scheme that samples from a quasiprobability distribution over stabiliser states [3, 33, 38]. These algorithms estimate the expectation value of a Pauli observable after a stabiliser channel is applied to a *non*-stabiliser input state. The minimum number of samples required to achieve some stated accuracy scales with $\mathcal{R}^2$.

The robustness of magic can be calculated using standard linear programming techniques [45] (for example using the MATLAB package CVX [46]). The naive formulation of the linear program is practical on a desktop computer for up to five qubits (the number of stabiliser states increases super-exponentially with n). It was recently shown by Heinrich and Gross [47] that when states possess certain symmetries, the original optimisation problem can be mapped to a more tractable one, so that the robustness of magic can be calculated for up to 10 copies of a state.

The framework naturally extends to a subclass of non-stabiliser circuits: those that may be implemented by deterministic state injection [38], including all gates from the third level of the Clifford hierarchy (Figure 2). The canonical example is the T-gate, $T = \mathrm{diag}(1, e^{i\pi/4})$, which can
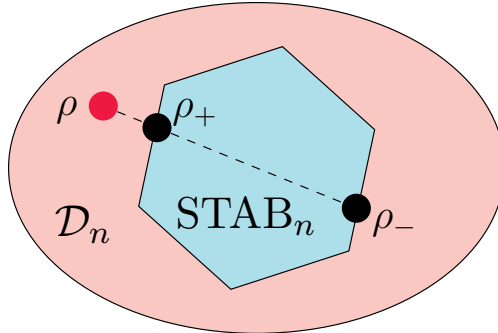
FIG. 1. Schematic illustration of a density matrix $\rho \in \mathcal{D}_n$ decomposed as an affine combination of elements from the stabiliser polytope STAB_n.
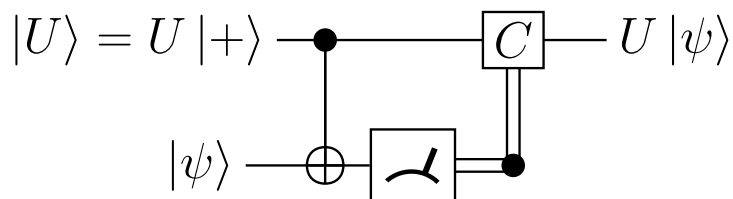

FIG. 2. State injection gadget. A resource state $|U\rangle$ is consumed in order to implement the corresponding gate U. A Clifford correction C is applied to qubit 1 conditioned on the outcome of a Pauli measurement on qubit 2. A single-qubit diagonal gate is shown, but the scheme can be generalised to all multi-qubit gates from the third level of the Clifford hierarchy.

be implemented by consuming so-called magic states as a resource [7]. The classical simulation overhead for implementing a gate is then the robustness of magic for the consumed resource state. Not all non-stabiliser operations can be implemented in this way, however.

Informally we say that an operation is stabiliser-preserving if it always maps stabiliser states to stabiliser states. To make this precise, define $\mathrm{SP}_{n,m}$ to be the set of n-qubit operations $\mathcal{E}$ such that $(\mathcal{E} \otimes \mathbb{1}_m)\sigma \in \mathrm{STAB}_{n+m}$ for all $\sigma \in \mathrm{STAB}_{n+m}$, where $\mathbb{1}_m$ is the identity map for an m-qubit Hilbert space. The set $\mathrm{SP}_{n,0}$ is then the set of channels that map n-qubit stabiliser states to n-qubit stabiliser states. We say a channel is "completely" stabiliser-preserving if $\mathcal{E} \in \mathrm{SP}_{n,m}$ for all m.

3. OVERVIEW OF MAIN RESULTS

Our first result is a characterisation of the class of completely stabiliser-preserving operations, making use of the well-known Choi-Jamiołkowski isomorphism [48–50]. Throughout this paper we will consider Hilbert spaces comprised of n-qubit subspaces, eg. $\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$. Where it is unclear from the context, we use superscripts to indicate which subspace each operator belongs to. For the maximally entangled state $|\Omega_n\rangle$, as defined in Theorem 3.1 below, we use the superscript $A|B$ to indicate the partition across which the state is entangled.

Theorem 3.1 (Completely stabiliser-preserving operations). *Given an n-qubit CPTP channel $\mathcal{E}$, for all $m > 0$, $\mathcal{E} \in \mathrm{SP}_{n,n+m}$ if and only if $\mathcal{E} \in \mathrm{SP}_{n,n}$. Furthermore, $\mathcal{E} \in \mathrm{SP}_{n,n}$ if and only if the Choi state*

$$\Phi_{\mathcal{E}}^{AB} = \left(\mathcal{E}^A \otimes \mathbb{1}_n\right)|\Omega_n\rangle\langle\Omega_n|^{A|B}, \quad where \quad |\Omega_n\rangle^{A|B} = \frac{1}{\sqrt{2^n}}\sum_{j=0}^{2^n-1}|j\rangle^A \otimes |j\rangle^B, \tag{3}$$

is a stabiliser state. Here, $|j\rangle$ are the n-qubit computational basis states.

We prove this in section 4 . We take this to be the set of free operations in our resource theories.

Our first new monotone is channel robustness $\mathcal{R}_*$. For an n-qubit CPTP channel $\mathcal{E}$ this is defined as:

$$\mathcal{R}_*(\mathcal{E}) = \min_{\Lambda_\pm \in \mathrm{SP}_{n,n} \cap \mathrm{CPTP}} \{2p + 1 : (1+p)\Lambda_+ - p\Lambda_- = \mathcal{E}, p \geq 0\}, \tag{4}$$

where $\Lambda_\pm$ are completely stabiliser-preserving and CPTP maps. To fully enumerate this class of maps, we notice that the associated Choi state must satisfy two conditions: (i) $\Phi_{\mathcal{E}}^{AB}$ is a stabiliser state, and (ii) $\Phi_{\mathcal{E}}^{AB}$ satisfies the trace-preservation condition $\mathrm{Tr}_A(\Phi_{\mathcal{E}}) = \frac{\mathbb{1}_n}{2^n}$. We can therefore write:

$$\mathcal{R}_*(\mathcal{E}) = \min_{\rho_\pm \in \mathrm{STAB}_{2n}} \left\{2p + 1 : (1+p)\rho_+ - p\rho_- = \Phi_{\mathcal{E}}^{AB}, p \geq 0, \mathrm{Tr}_A(\rho_\pm) = \frac{\mathbb{1}_n}{2^n}\right\}. \tag{5}$$

This can now be calculated by linear program given access to a list of all stabiliser states (see Appendix C and the code repository Ref. [44]). Channel robustness satisfies the following:

1. *Faithfulness*: If $\mathcal{E}$ is a CPTP channel, then $\mathcal{R}_*(\mathcal{E}) = 1$ if $\mathcal{E}$ is completely stabiliser-preserving and strictly larger than 1 otherwise;

2. *Convexity*: $\mathcal{R}_*\left(\sum_j q_j \mathcal{E}_j\right) \leq \sum_j |q_j|\mathcal{R}_*(\mathcal{E}_j)$;

3. *Submultiplicativity under composition*: $\mathcal{R}_*(\mathcal{E}_2 \circ \mathcal{E}_1) \leq \mathcal{R}_*(\mathcal{E}_2)\mathcal{R}_*(\mathcal{E}_1)$;

4. *Submultiplicativity under tensor product*: $\mathcal{R}_*\left(\mathcal{E}^A \otimes \mathcal{E}'^B\right) \leq \mathcal{R}_*\left(\mathcal{E}^A\right)\mathcal{R}_*\left(\mathcal{E}'^B\right)$.

As a special case, if Λ is a CPTP stabiliser channel, then

$$\mathcal{R}_*(\Lambda \circ \mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})\mathcal{R}_*(\Lambda) = \mathcal{R}_*(\mathcal{E}), \tag{6}$$

and similarly $\mathcal{R}_*(\mathcal{E} \circ \Lambda) \leq \mathcal{R}_*(\mathcal{E})$. This combines submultiplicativity under composition and faithfulness, to show that $\mathcal{R}_*$ is suitably monotonically non-increasing under compositions with stabiliser channels. This is the sense in which channel robustness is a magic monotone for channels. We prove submultiplicativity in Section 5. For completeness we prove convexity and faithfulness in Appendix B.

The approach above is very close to the stabiliser decomposition of channels employed by Bennink et al. in Ref. [3]. The main difference is that Bennink et al. optimise their decomposition with respect to CPR, the set of Clifford unitaries supplemented by Pauli reset channels, rather than $\mathrm{SP}_{n,n}$. The set CPR turns out to be a strict subset of the stabiliser-preserving CPTP maps, so $\mathcal{R}_*$ is a lower bound to the ℓ_1-norm of any CPR decomposition (though the bound is tight in many cases). Just as the ℓ_1-norm in Bennink et al. quantifies the sample complexity of a classical simulation algorithm, we can construct a related algorithm where the runtime depends on $\mathcal{R}_*$ in a similar way. We give the details of this algorithm in Section 7 A.

Before proceeding, let us reflect on the condition $\mathrm{Tr}_A(\rho_\pm) = \mathbb{1}_n/2^n$ that enforces that the corresponding channels $\mathcal{E}_\pm$ are trace-preserving. Dropping this condition would instead lead to

$\mathcal{R}(\Phi_{\mathcal{E}})$, the robustness of the Choi state. For gates from the third level of the Clifford hierarchy, deterministic state injection is always possible, and hence the resource cost of the gate U can be equated with the robustness of magic of the corresponding resource state. These resource states can always (by Clifford-equivalence) be taken to have the form $|U\rangle\langle U| = (U_A \otimes \mathbb{1}_B) |\Omega\rangle\langle\Omega|$. This is precisely the Choi state, so it is natural to ask if $\mathcal{R}(\Phi_{\mathcal{E}})$ also quantifies non-stabiliserness for more general channels. We find that $\mathcal{R}(\Phi_{\mathcal{E}})$ exhibits faithfulness, convexity and submultiplicativity under tensor product, but lacks submultiplicativity under composition. This arises from the fact that the decomposition of the Choi state corresponds to a decomposition of the channel into maps that are not necessarily trace-preserving. See Appendix A for a counterexample. Despite this shortcoming, we will see that $\mathcal{R}(\Phi_{\mathcal{E}})$ is a useful quantity to compare to more well-behaved measures. Moreover, given that $\mathcal{R}(\Phi_{\mathcal{E}})$ does give the resource cost for third level Clifford hierarchy gates, for consistency of the framework it is reasonable to require that our new monotones should be equal to $\mathcal{R}(\Phi_{\mathcal{E}})$ for this restricted class of gates. We will show later that our monotones do have this property.

Our second new monotone is the *magic capacity*. Given an n-qubit channel $\mathcal{E}$, it is natural to consider the largest possible increase in robustness of magic, over any possible input state. By analogy with the resource theories of entanglement [51] and coherence [24], we define the magic capacity as:

$$\mathcal{C}(\mathcal{E}) = \max_{|\phi\rangle \in \text{STAB}_{2n}} \mathcal{R}[(\mathcal{E} \otimes \mathbb{1}_n) |\phi\rangle\langle\phi|].$$

Note that the definition of capacity involves forming a tensor product of an n-qubit channel $\mathcal{E}$ with the n-qubit identity. This is necessary because there exist n-qubit channels that generate their maximum robustness when applied to part of an m-qubit state, where $m > n$. Nevertheless, the n-qubit identity suffices for our definition; This is a consequence of Lemma 4.1 in Section 4. The capacity has the following useful properties:

1. *Faithfulness*: If $\mathcal{E}$ is a completely positive, trace-preserving (CPTP) channel, then $\mathcal{C} = 1$ if $\mathcal{E}$ is stabiliser-preserving (SP), and strictly larger than 1 otherwise;

2. *Convexity*: $\mathcal{C}\left(\sum_j q_j \mathcal{E}_j\right) \leq \sum_j |q_j| \mathcal{C}(\mathcal{E}_j)$;

3. *Submultiplicativity under composition*: $\mathcal{C}(\mathcal{E}_1 \circ \mathcal{E}_2) \leq \mathcal{C}(\mathcal{E}_1)\mathcal{C}(\mathcal{E}_2)$;

4. *Submultiplicativity under tensor product*: $\mathcal{C}\left(\mathcal{E}^A \otimes \mathcal{E}'^B\right) \leq \mathcal{C}\left(\mathcal{E}^A\right)\mathcal{C}\left(\mathcal{E}'^B\right)$;

5. *Maximum increase in robustness*: $\frac{\mathcal{R}[(\mathcal{E} \otimes \mathbb{1})\rho]}{\mathcal{R}(\rho)} \leq \mathcal{C}(\mathcal{E})$, $\forall \rho$.

In Section 6 we prove properties 3-5, (convexity and faithfulness are shown in Appendix D). We will also prove the following theorem relating magic capacity to channel robustness and $\mathcal{R}(\Phi_{\mathcal{E}})$.

Theorem 3.2 (Sandwich Theorem). *For any CPTP map $\mathcal{E}$, the following inequalities hold:*

$$\mathcal{R}(\Phi_{\mathcal{E}}) \leq \mathcal{C}(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E}). \tag{7}$$

Moreover, if the unitary operation $\mathcal{U}$ is in the third level of the Clifford hierarchy, then we have equality:

$$\mathcal{R}(\Phi_{\mathcal{U}}) = \mathcal{C}(\mathcal{U}) = \mathcal{R}_*(\mathcal{U}). \tag{8}$$

We are interested in whether or not these inequalities are tight for more general operations. In Table I we summarise numerical results for a selection of diagonal gates. The results for these gates are presented in full in Section 8.

n	2	3	4	5
Multicontrol gates, $t = 0$	$\mathcal{R}_\Phi = \mathcal{C} = \mathcal{R}_*$	$\mathcal{R}_\Phi = \mathcal{C} = \mathcal{R}_*$	$\mathcal{R}_\Phi = \mathcal{C} = \mathcal{R}_*$	$\mathcal{R}_\Phi = \mathcal{C} < \mathcal{R}$
Multicontrol gates, $t \geq 1$	$\mathcal{R}_\Phi = \mathcal{C} = \mathcal{R}_*$	$\mathcal{R}_\Phi = \mathcal{C} = \mathcal{R}_*$	$\mathcal{R}_\Phi = \mathcal{C} < \mathcal{R}_*$	$\mathcal{R}_\Phi < \mathcal{C} < \mathcal{R}$
Random phase gates	$\mathcal{R}_\Phi = \mathcal{C} = \mathcal{R}_*$	$\mathcal{R}_\Phi = \mathcal{C} \leq \mathcal{R}_*$	$\mathcal{R}_\Phi = \mathcal{C} \leq \mathcal{R}_*$	-

TABLE I. Tightness of bound given by Theorem 3.2, as determined by numerical estimation of diagonal gates, where $\mathcal{R}_\Phi$ is the robustness of the Choi state, $\mathcal{C}$ is the magic capacity, $\mathcal{R}_*$ is the trace-preserving variant of $\mathcal{R}_\Phi$. Here an equality indicates that in all cases investigated, values calculated were equal up to the precision of the solver. Multicontrol phase gates are taken to be those represented by unitaries of the form $\mathrm{diag}(1, \ldots, 1, \exp[i\pi/2^t])$.

The magic capacity also quantifies the sample complexity for a Monte Carlo-type classical simulation algorithm, presented in Section 7 B. This differs from previous algorithms such as Bennink et al. [3] in that a convex optimisation must be solved at each step. While this results in an increase in runtime per sample, it can be the case that $\mathcal{C}(\mathcal{E}) \ll \mathcal{R}_*(\mathcal{E})$, which can lead to an improvement in sample complexity over the algorithm of Section 7 A.

4. COMPLETELY STABILISER-PRESERVING OPERATIONS

In this section, we justify setting $\mathrm{SP}_{n,n}$ as the class of free operations. We begin with an example channel $\mathcal{E} \in \mathrm{SP}_{n,0}$ that fails to be stabiliser-preserving when acting on part of a larger system. Consider the single-qubit channel $\mathcal{E}_T$ defined by the Kraus operators $\{|0\rangle\langle T|, |1\rangle\langle T_\perp|\}$, where $|T\rangle = T|+\rangle$ and $|T_\perp\rangle = T|-\rangle$. Clearly, applied to any single-qubit state, the output will be some probabilistic mixture of $|0\rangle$ and $|1\rangle$, and so must have $\mathcal{R} = 1$, so $\mathcal{E}_T \in \mathrm{SP}_{n,0}$. But if $\mathcal{E}_T$ is applied to one qubit in a Bell pair, we obtain:

$$(\mathcal{E}_T \otimes \mathbb{1})|\Phi_+\rangle\langle\Phi_+| = \frac{1}{2}(|0T^*\rangle\langle 0T^*| + |1T_\perp^*\rangle\langle 1T_\perp^*|), \tag{9}$$

where $|T^*\rangle = T^\dagger|+\rangle$, $|T_\perp^*\rangle = T^\dagger|-\rangle$. From this output state, we can deterministically recover a pure magic state on qubit 2 using only stabiliser operations, by making a Z-measurement on qubit 1 and then performing a rotation on qubit 2 conditioned on the outcome. The output state has robustness $\mathcal{R}((\mathcal{E}_T \otimes \mathbb{1})|\Phi_+\rangle\langle\Phi_+|) = \mathcal{R}(|T\rangle) = \sqrt{2}$.

So, there exist channels where $\mathcal{E} \in \mathrm{SP}_{n,m}$ but $\mathcal{E} \notin \mathrm{SP}_{n,m+1}$. To call a channel *completely stabiliser-preserving*, then, we need to be sure $\mathcal{E} \otimes \mathbb{1}_m$ remains stabiliser-preserving for all $m > 0$. We now show we only need tensor with the identity of the same dimension as the original channel.

Lemma 4.1 (Maximum robustness achieved on $2n$ qubits)**.** *Let $\mathcal{E}$ be an n-qubit quantum channel. Then for $m > 0$, for any $|\phi\rangle \in \mathrm{STAB}_{2n+m}$, there exists some state $|\psi\rangle \in \mathrm{STAB}_{2n}$ such that:*

$$\mathcal{R}\left[\left(\mathcal{E}^A \otimes \mathbb{1}_{n+m}\right)|\phi\rangle\langle\phi|^{AB}\right] = \mathcal{R}\left[\left(\mathcal{E}^A \otimes \mathbb{1}_n\right)|\psi\rangle\langle\psi|^{AB'}\right]. \tag{10}$$

Proof. Consider a $(2n + m)$-qubit stabiliser state $|\phi\rangle$, with partition $A|B$ between the first n and last $n+m$ qubits. Ref. [52] shows that the state $|\phi\rangle^{AB}$ is local Clifford-equivalent to p independent Bell pairs entangled across the partition $A|B$ (here "local" means with respect to the bipartition rather than per qubit). Since there are n qubits in partition A, p is at most n. Let $B'|B''$ be a partition of B into n and m qubits. Then by local permutation of qubits within B, we can take these $p \leq n$ Bell pairs to be entangled across $A|B'$. So we have:

$$|\phi\rangle^{AB} = \left(\mathbb{1}_n \otimes U^B\right)|\psi\rangle^{AB'}|\psi'\rangle^{B''}, \tag{11}$$

where U^B is a Clifford operation, $|\psi\rangle^{AB'} \in \mathrm{STAB}_{2n}$ and $|\psi'\rangle^{B''} \in \mathrm{STAB}_m$. So writing the channel corresponding to U^B as $\mathcal{U}^B$, for any $\mathcal{E}$ on n qubits, we know that:

$$\mathcal{R}\Big[\big(\mathcal{E}^A \otimes \mathbb{1}_{n+m}\big)\,|\phi\rangle\langle\phi|^{AB}\Big] = \mathcal{R}\Big[\big(\mathbb{1}_n \otimes \mathcal{U}^B\big)\Big(\big(\mathcal{E}^A \otimes \mathbb{1}_n\big)\big(|\psi\rangle\langle\psi|^{AB'}\big) \otimes |\psi'\rangle\langle\psi'|^{B''}\Big)\Big]. \qquad (12)$$

Since $\mathbb{1}_n \otimes \mathcal{U}^B$ represents a (reversible) Clifford gate, by monotonicity of robustness of magic:

$$\mathcal{R}\Big[\big(\mathcal{E}^A \otimes \mathbb{1}_{n+m}\big)\,|\phi\rangle\langle\phi|^{AB}\Big] = \mathcal{R}\Big[\big(\mathcal{E}^A \otimes \mathbb{1}_n\big)\big(|\psi\rangle\langle\psi|^{AB'}\big) \otimes |\psi'\rangle\langle\psi'|^{B''}\Big] \qquad (13)$$

$$= \mathcal{R}\Big[\big(\mathcal{E}^A \otimes \mathbb{1}_n\big)\,|\psi\rangle\langle\psi|^{AB'}\Big], \qquad (14)$$

where in the last line we used the fact that $|\psi'\rangle^{B''}$ is a stabiliser state, and hence does not contribute to the robustness. The state $|\psi\rangle\langle\psi|^{AB'}$ is a $2n$-qubit state, so this proves the result. $\qquad\square$

This lemma allows us to prove the first claim of Theorem 3.1, which says that $\mathcal{E}$ is completely stabiliser-preserving if and only if $\mathcal{E} \in \mathrm{SP}_{n,n}$. The inclusion $\mathrm{SP}_{n,n} \subseteq \mathrm{SP}_{n,n+m}$ is immediate since the stabiliser states are preserved under tracing out of auxiliary systems. The interesting inclusion is $\mathrm{SP}_{n,n+m} \subseteq \mathrm{SP}_{n,n}$. Suppose that $\mathcal{E} \in \mathrm{SP}_{n,n}$ and consider any $\sigma \in \mathrm{STAB}_{2n+m}$. By Lemma 4.1 there exists some stabiliser state $\sigma' \in \mathrm{STAB}_{2n}$ such that $\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_{n+m})\sigma) = \mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)\sigma')$. But if $\mathcal{E} \in \mathrm{SP}_{n,n}$, then $(\mathcal{E} \otimes \mathbb{1}_n)\sigma'$ is a stabiliser state, so the robustness is equal to 1. By the faithfulness of robustness of magic, $(\mathcal{E} \otimes \mathbb{1}_{n+m})\sigma$ is a stabiliser state. Therefore, $\mathcal{E} \in \mathrm{SP}_{n,n}$ implies $\mathcal{E} \in \mathrm{SP}_{n,n+m}$. Next, we discuss a straightforward test for membership of this set, which does not require mechanically checking all possible input stabiliser states.

We can associate every n-qubit channel $\mathcal{E}$ with a unique density operator on $2n$ qubits [48–50]:

$$\Phi_{\mathcal{E}}^{AB} = \big(\mathcal{E}^A \otimes \mathbb{1}^B\big)\,|\Omega_n\rangle\langle\Omega_n|^{A|B}, \quad \text{where} \quad |\Omega_n\rangle^{A|B} = \frac{1}{\sqrt{2^n}}\sum_{j=0}^{2^n-1}|j\rangle^A \otimes |j\rangle^B. \qquad (15)$$

Here $|j\rangle$ label the computational basis states. We will also use the following property:

$$\mathrm{Tr}[A\mathcal{E}(\rho)] = 2^n\,\mathrm{Tr}\big[\Phi_{\mathcal{E}}(A \otimes \rho^T)\big], \quad \forall \rho, A. \qquad (16)$$

Consider the robustness of magic of the Choi state, $\mathcal{R}(\Phi_{\mathcal{E}})$. We mentioned earlier that $\mathcal{R}(\Phi_{\mathcal{E}})$ quantifies simulation cost for gates from the third level of the Clifford hierarchy. This motivates us to consider its properties for more general operations, and it turns out that $\mathcal{R}(\Phi_{\mathcal{E}})$ gives us our first criterion for completely stabiliser-preserving channels.

Lemma 4.2 (Faithfulness of robustness of the Choi state). *Consider the n-qubit CPTP channel $\mathcal{E}$. If $\mathcal{E} \in \mathrm{SP}_{n,n}$, then $\mathcal{R}(\Phi_{\mathcal{E}}) = 1$. Otherwise, $\mathcal{R}(\Phi_{\mathcal{E}}) > 1$.*

Proof. The fact that $\mathcal{E} \in \mathrm{SP}_{n,n}$ implies $\mathcal{R}(\Phi_{\mathcal{E}}) = 1$ is easy to see. Since $|\Omega_n\rangle\langle\Omega_n|$ is itself a $2n$-qubit stabiliser state, $\mathcal{E} \in \mathrm{SP}_{n,n}$ guarantees that $\Phi_{\mathcal{E}}$ is a stabiliser state, so must have robustness 1. The implication in the other direction is less obvious; one might imagine there perhaps exist maps that send $|\Omega_n\rangle\langle\Omega_n|$ in particular to a stabiliser state, but are not stabiliser-preserving in general. We show that this is not the case using an argument based on witnesses for non-stabiliser states, in part inspired by the conditions for free operations (SPO) given by Ahmadi et al. [31] for odd prime dimension qudits and for the single qubit case. The criteria for SPO were based on a class of witness defined by phase point operators. Here we instead consider the following family of witnesses for n qubits. We say that W_n is a good witness for n-qubit non-stabiliser states if:

$$\mathrm{Tr}(W_n\sigma) \leq 0, \quad \forall \sigma \in \mathrm{STAB}_n. \qquad (17)$$

The hyperplane separation theorem [45] guarantees that such witnesses exist and can be constructed for any non-stabiliser state ρ. That is, for any $\rho \notin \text{STAB}_n$, there always exists an operator W_ρ such that $\text{Tr}(W_\rho \rho) > 0$ and yet is a good n-qubit witness as defined above.

We first show that for any good n-qubit witness W_n, the operator $W_n \otimes |\phi\rangle\langle\phi|$, where $|\phi\rangle \in \text{STAB}_m$, is a good witness for $(n+m)$-qubit non-stabiliser states. For any $\sigma \in \text{STAB}_{n+m}$:

$$\text{Tr}[(W_n \otimes |\phi\rangle\langle\phi|)\sigma] = \text{Tr}[(\mathbb{1}_n \otimes |\phi\rangle\langle\phi|)(W_n \otimes \mathbb{1}_m)\sigma] = \text{Tr}[(W_n \otimes \mathbb{1}_m)\widetilde{\sigma}], \tag{18}$$

so that $\widetilde{\sigma} = (\mathbb{1}_n \otimes |\phi\rangle\langle\phi|)\sigma(\mathbb{1}_n \otimes |\phi\rangle\langle\phi|)$, where we used cyclicity of the trace and the fact that $|\phi\rangle\langle\phi|$ is a projector. If $\widetilde{\sigma} = 0$ then the inequality (17) is trivially always satisfied by $W_n \otimes |\phi\rangle\langle\phi|$. Otherwise, $\widetilde{\sigma}$ is a stabiliser state (non-normalised) and so too is $\text{Tr}_B[\widetilde{\sigma}]$. Then:

$$\text{Tr}[(W_n \otimes \mathbb{1}_m)\widetilde{\sigma}] = \text{Tr}[W_n \text{Tr}_B(\widetilde{\sigma})] \leq 0. \tag{19}$$

The inequality follows because W_n is a good witness and $\text{Tr}_B[\widetilde{\sigma}]$ is a stabiliser state. Therefore, $W_n \otimes |\phi\rangle\langle\phi|$ is also a valid witness.

Now suppose $\mathcal{E} \notin \text{SP}_{n,n}$. Then there is some stabiliser state $|\phi\rangle \in \text{STAB}_{2n}$, such that $\rho' = (\mathcal{E} \otimes \mathbb{1}_n)|\phi\rangle\langle\phi| \notin \text{STAB}_{2n}$. By the hyperplane separation theorem, there exists a good $2n$-qubit witness $W_{\rho'}$ such that $\text{Tr}[W_{\rho'}\rho'] > 0$. Consider that the $4n$-qubit state $|\Omega_{2n}\rangle^{AA'|BB'}$ is unentangled across the partition $AB|A'B'$, so that we can write:

$$|\Omega_{2n}\rangle^{AA'|BB'} = \frac{1}{2^{2n}} \sum_{j,k} |j\rangle^A |k\rangle^{A'} \otimes |j\rangle^B |k\rangle^{B'} = |\Omega_n\rangle^{A|B} \otimes |\Omega_n\rangle^{A'|B'}, \tag{20}$$

taking care to note the permutation of subspaces. Therefore the Choi state for $(\mathcal{E} \otimes \mathbb{1}_n)$ is:

$$\Phi_{\mathcal{E}\otimes\mathbb{1}_n}^{AA'|BB'} = (\mathcal{E}^A \otimes \mathbb{1}_n^{A'}) \otimes \mathbb{1}_{2n}^{BB'}(|\Omega_{2n}\rangle\langle\Omega_{2n}|^{AA'|BB'}) = \Phi_{\mathcal{E}}^{AB} \otimes |\Omega_n\rangle\langle\Omega_n|^{A'|B'}. \tag{21}$$

We then use equation (16) to obtain:

$$0 < \frac{1}{2^n} \text{Tr}[W_{\rho'}\rho'] = \text{Tr}\left[\Phi_{\mathcal{E}\otimes\mathbb{1}_n}(W_{\rho'} \otimes |\phi'\rangle\langle\phi'|^T)\right], \quad \text{where} \quad |\phi'\rangle\langle\phi'|^T \in \text{STAB}_{2n}. \tag{22}$$

But $(W_{\rho'} \otimes |\phi'\rangle\langle\phi'|^T)$ is a good witness, so $\Phi_{\mathcal{E}} \otimes |\Omega_n\rangle\langle\Omega_n| \notin \text{STAB}_{4n}$ and therefore $\Phi_{\mathcal{E}}$ is a non-stabiliser state. So, by faithfulness of robustness of magic, if $\mathcal{E} \notin \text{SP}_{n,n}$ then $\mathcal{R}(\Phi_{\mathcal{E}}) > 1$. $\qquad\square$

Combined the above two lemmas provide a proof of both claims given in Theorem 3.1. This does not mean the robustness of the Choi state is a reliable monotone, since despite being faithful it fails to be submultiplicative under composition (see Appendix A). Rather, we use the faithfulness of the Choi state as a tool to give an alternative definition of the channel robustness as captured by Eq. (5).

5. CHANNEL ROBUSTNESS

A natural extension of the algorithm of Bennink et al. [3] is to replace CPR (the set of Clifford gates and Pauli reset channels) with $\text{SP}_{n,n}$. We therefore define the *channel robustness* as:

$$\mathcal{R}_*(\mathcal{E}) = \min_{\Lambda_\pm \in \text{SP}_{n,n}} \{2p+1 : (1+p)\Lambda_+ - p\Lambda_- = \mathcal{E}, p \geq 0\}. \tag{23}$$

To calculate this in practice, we decompose the Choi state $\Phi_{\mathcal{E}}$ as per equation (5), adapting the robustness of magic optimisation problem from Ref. [38]. The details are given in Appendix C. We

also note that for diagonal channels $\mathcal{E}$, the problem is equivalent to a decomposition of the state $\mathcal{E}(|+\rangle\langle+|^{\otimes n})$ as:

$$\mathcal{E}(|+\rangle\langle+|^{\otimes n}) = (1+p)\rho_+ - p\rho_-, \text{ where } \rho_\pm \in \text{STAB}_n, \text{ and } \langle x| \rho_\pm |x\rangle = \frac{1}{2^n}, \ \forall x. \tag{24}$$

Here, the condition on the partial trace for the general case is replaced by the requirement that all diagonal elements of the states $\rho_\pm$ are equal to $1/2^n$. In the special case where $\rho_\pm$ are pure, this implies they are diagonal Clifford-equivalent to graph states, though more generally they may be mixtures of states that individually do not have full support in the standard basis. In practical terms, this reduction to an n-qubit problem confers a significant advantage, since the number of stabiliser states (which form the extreme points of the linear programming problem) grows super-exponentially with n. Full technical details of this simplification are given in Appendix E. We now return to consider the properties of channel robustness.

The channel robustness is convex and faithful with these properties inherited from the robustness of magic (see Appendix B for details). Here we discuss additional properties.

Submultiplicativity under composition: $\mathcal{R}_*(\mathcal{E}_2 \circ \mathcal{E}_1) \leq \mathcal{R}_*(\mathcal{E}_1)\mathcal{R}_*(\mathcal{E}_2)$. The channels $\mathcal{E}_1$ and $\mathcal{E}_2$ will have an optimal decomposition:

$$\mathcal{E}_j = (1+p_j)\Lambda_{j,+} - p_j\Lambda_{j,-}, \tag{25}$$

where $\mathcal{R}_*(\mathcal{E}_j) = 1 + 2p_j$ and $\Lambda_{j,\pm}$ are CPTP maps and completely stabiliser preserving. Using these decompositions, we obtain that

$$\mathcal{E}_2 \circ \mathcal{E}_1 = (1+q)\Lambda'_+ - q\Lambda'_-, \tag{26}$$

where

$$\Lambda'_+ = (1+q)^{-1}[(1+p_2)(1+p_1)\Lambda_{2,+} \circ \Lambda_{1,+} + p_2 p_1 \Lambda_{2,-} \circ \Lambda_{1,-}], \tag{27}$$

$$\Lambda'_- = q^{-1}[p_2(1+p_1)\Lambda_{2,-} \circ \Lambda_{1,+} + (1+p_2)p_1\Lambda_{2,+} \circ \Lambda_{1,-}], \tag{28}$$

$$q = p_1 + p_2 + 2p_1 p_2 \tag{29}$$

The set of CPTP completely stabiliser preserving channels is closed under composition and convex, so both $\Lambda'_\pm$ are in this set. Therefore, we have a valid decomposition for $\mathcal{E}_2 \circ \mathcal{E}_1$ that entails $\mathcal{R}_*(\mathcal{E}_2 \circ \mathcal{E}_1) \leq 1 + 2q$. One finds

$$1 + 2q = (1+2p_1)(1+2p_2) = \mathcal{R}_*(\mathcal{E}_1)\mathcal{R}_*(\mathcal{E}_2), \tag{30}$$

which completes the proof.

Submultiplicativity under tensor product: $\mathcal{R}_*(\mathcal{E}^A \otimes \mathcal{E}'^B) \leq \mathcal{R}_*(\mathcal{E}^A)\mathcal{R}_*(\mathcal{E}'^B)$. We treat tensor product as a special case of composition. For n-qubit $\mathcal{E}_A$ and m-qubit $\mathcal{E}'_B$:

$$\mathcal{R}_*(\mathcal{E}^A \otimes \mathcal{E}'^B) \leq \mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}_m^B)\mathcal{R}_*(\mathbb{1}_n^A \otimes \mathcal{E}'^B). \tag{31}$$

To complete the proof we will confirm that

$$\mathcal{R}_*(\mathbb{1}_n^A \otimes \mathcal{E}^B) = \mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}_n^B) = \mathcal{R}_*(\mathcal{E}). \tag{32}$$

As noted earlier, we can write $|\Omega_{n+m}\rangle^{AA'|BB'} = |\Omega_n\rangle^{A|B} \otimes |\Omega_m\rangle^{A'|B'}$, so that the Choi state for

$\mathcal{E}^A \otimes \mathbb{1}_m$ is given by:

$$
\begin{aligned}
\Phi_{\mathcal{E}\otimes\mathbb{1}_m}^{AA'|BB'} &= \left(\mathcal{E}^A \otimes \mathbb{1}_m^{A'} \otimes \mathbb{1}_{n+m}^{BB'} \right) |\Omega_{n+m}\rangle\langle\Omega_{n+m}|^{AA'|BB'} \\
&= \left(\mathcal{E}^A \otimes \mathbb{1}_n^B \right) |\Omega_n\rangle\langle\Omega_n|^{A|B} \otimes |\Omega_m\rangle\langle\Omega_m|^{A'|B'} \\
&= \Phi_{\mathcal{E}}^{AB} \otimes |\Omega_m\rangle\langle\Omega_m|^{A'|B'} .
\end{aligned}
\tag{33}
$$

The state $\Phi_{\mathcal{E}}$ will have some optimal decomposition $\Phi_{\mathcal{E}} = (1+p)\rho_+ - p\rho_-$, with channel robustness $\mathcal{R}_*(\mathcal{E}_A) = 1 + 2p$, so that:

$$
\Phi_{\mathcal{E}\otimes\mathbb{1}_m}^{AA'|BB'} = (1+p)\rho_+^{AB} \otimes |\Omega_m\rangle\langle\Omega_m|^{A'|B'} - p\rho_-^{AB} \otimes |\Omega_m\rangle\langle\Omega_m|^{A'|B'} .
\tag{34}
$$

This is a valid, not necessarily optimal, stabiliser decomposition satisfying the trace condition, so we have

$$
\mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}^{A'}) \leq \mathcal{R}_*(\mathcal{E}^A).
\tag{35}
$$

This is enough to show submultiplicativity; for completeness, in Appendix B we will also show $\mathcal{R}_*(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E} \otimes \mathbb{1})$ so that in fact we have equality.

6. MAGIC CAPACITY

A. Properties

We now turn to our second monotone, which quantifies the capacity of a channel to generate magic. Recall:

$$
\mathcal{C}(\mathcal{E}) = \max_{|\phi\rangle\in\text{STAB}_{2n}} \mathcal{R}[(\mathcal{E} \otimes \mathbb{1}_n) |\phi\rangle\langle\phi|],
\tag{36}
$$

where $\mathcal{R}$ is the robustness of magic. Notice that we only need optimise over the pure stabiliser states. For mixed states or even non-stabiliser states, the capacity still captures the possible increase in robustness of magic by virtue of the **maximum increase in robustness** property:

$$
\frac{\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)\rho)}{\mathcal{R}(\rho)} \leq \mathcal{C}(\mathcal{E}).
\tag{37}
$$

Here we prove this property, using similar arguments to those deployed in [51]. Consider an n-qubit channel $\mathcal{E}$. Any $2n$-qubit input state ρ will have an optimal stabiliser state decomposition $\rho = \sum_j q_j |\phi_j\rangle\langle\phi_j|$, where $\sum_j q_j = 1$, and such that $\mathcal{R}(\rho) = \sum_j |q_j|$. By linearity we have:

$$
(\mathcal{E} \otimes \mathbb{1}_n)\rho = \sum_j q_j(\mathcal{E} \otimes \mathbb{1}_n) |\phi_j\rangle\langle\phi_j| .
\tag{38}
$$

By convexity of robustness of magic, we then have:

$$
\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)\rho) \leq \sum_j |q_j| \mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n) |\phi_j\rangle\langle\phi_j|).
\tag{39}
$$

The optimal pure stabiliser state $|\phi_*\rangle$, satisfies:

$$
\mathcal{C}(\mathcal{E}) = \mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n) |\phi_*\rangle\langle\phi_*|) \geq \mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n) |\phi_j\rangle\langle\phi_j|)
$$

for any j. So we have:

$$\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)\rho) \leq \mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n) |\phi_*\rangle\langle\phi_*|) \sum_j |q_j| = \mathcal{C}(\mathcal{E})\mathcal{R}(\rho). \tag{40}$$

Rearranging we obtain inequality (37).

Submultiplicativity under composition: $\mathcal{C}(\mathcal{E}_1 \circ \mathcal{E}_2) \leq \mathcal{C}(\mathcal{E}_1)\mathcal{C}(\mathcal{E}_2)$. Take the composition of two linear maps $\mathcal{E}_1$ and $\mathcal{E}_2$. There exists some stabiliser state $\rho_* = |\phi_*\rangle\langle\phi_*|$ that achieves the optimal robustness:

$$\mathcal{C}(\mathcal{E}_2 \circ \mathcal{E}_1) = \mathcal{R}([(\mathcal{E}_2 \circ \mathcal{E}_1) \otimes \mathbb{1}_n]\rho_*) = \mathcal{R}((\mathcal{E}_2 \otimes \mathbb{1}_n) \circ (\mathcal{E}_1 \otimes \mathbb{1}_n)\rho_*). \tag{41}$$

The operator $(\mathcal{E}_1 \otimes \mathbb{1}_n)\rho_*$ will have some optimal decomposition $(\mathcal{E}_1 \otimes \mathbb{1}_n)\rho_* = \sum_k q_{1k} |\phi_k\rangle\langle\phi_k|$ such that $\mathcal{R}((\mathcal{E}_1 \otimes \mathbb{1}_n)\rho_*) = \sum_k |q_{1k}|$. So by linearity:

$$(\mathcal{E}_2 \otimes \mathbb{1}_n) \circ (\mathcal{E}_1 \otimes \mathbb{1}_n)[\rho_*] = (\mathcal{E}_2 \otimes \mathbb{1}_n)\left[\sum_k q_{1k} |\phi_k\rangle\langle\phi_k|\right] = \sum_k q_{1k}(\mathcal{E}_2 \otimes \mathbb{1}_n) |\phi_k\rangle\langle\phi_k|. \tag{42}$$

Then by convexity of robustness of magic:

$$\mathcal{R}((\mathcal{E}_2 \otimes \mathbb{1}_n) \circ (\mathcal{E}_1 \otimes \mathbb{1}_n)[\rho_*]) = \mathcal{R}\left(\sum_k q_{1k}(\mathcal{E}_2 \otimes \mathbb{1}_n) |\phi_k\rangle\langle\phi_k|\right) \tag{43}$$

$$\leq \sum_k |q_{1k}|\mathcal{R}((\mathcal{E}_2 \otimes \mathbb{1}_n) |\phi_k\rangle\langle\phi_k|) \tag{44}$$

$$\leq \sum_k |q_{1k}|\mathcal{C}(\mathcal{E}_2) \tag{45}$$

$$= \mathcal{R}((\mathcal{E}_1 \otimes \mathbb{1}_n)\rho_*)\mathcal{C}(\mathcal{E}_2), \tag{46}$$

where to go from (44) to (45) we used the fact that since $|\phi_k\rangle$ are stabiliser states, $\mathcal{R}((\mathcal{E}_2 \otimes \mathbb{1}) |\phi_k\rangle\langle\phi_k|)$ can be no larger than $\mathcal{C}(\mathcal{E}_2)$. Finally, using the fact that $\mathcal{R}((\mathcal{E}_1 \otimes \mathbb{1})\rho_*) \leq \mathcal{C}(\mathcal{E}_1)$, we have $\mathcal{C}(\mathcal{E}_2 \circ \mathcal{E}_1) \leq \mathcal{C}(\mathcal{E}_2)\mathcal{C}(\mathcal{E}_1)$, completing the proof.

Submultiplicativity under tensor product: $\mathcal{C}(\mathcal{E}^A \otimes \mathcal{E}'^B) \leq \mathcal{C}(\mathcal{E}^A)\mathcal{C}(\mathcal{E}'^B)$. This follows directly from submultiplicativity under composition, since

$$\mathcal{C}(\mathcal{E}^A \otimes \mathcal{E}'^B) = \mathcal{C}((\mathcal{E}^A \otimes \mathbb{1}_m^B) \circ (\mathbb{1}_n^A \otimes \mathcal{E}'^B)) \leq \mathcal{C}(\mathcal{E}^A \otimes \mathbb{1}_m^B)\mathcal{C}(\mathbb{1}_n^A \otimes \mathcal{E}'^B). \tag{47}$$

We saw in Section 4 that any gains in robustness achievable by tensoring $\mathcal{E}_A$ with the identity and acting on a larger Hilbert space are already taken care of by the $\otimes \mathbb{1}_n$ in the definition (36), so that $\mathcal{C}(\mathcal{E}^A \otimes \mathbb{1}_m^B) = \mathcal{C}(\mathcal{E}^A)$ and $\mathcal{C}(\mathbb{1}_n^A \otimes \mathcal{E}'^B) = \mathcal{C}(\mathcal{E}'^B)$. Substituting this into inequality (47) gives the desired result.

B.   Sandwich theorem

We will now prove Theorem 3.2, which stated that $\mathcal{R}(\Phi_{\mathcal{E}}) \leq \mathcal{C}(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})$, for any CPTP channel $\mathcal{E}$, and that $\mathcal{R}(\Phi_{\mathcal{U}}) = \mathcal{C}(\mathcal{U}) = \mathcal{R}_*(\mathcal{U})$ for any unitary operation $\mathcal{U}$ from the third level of the Clifford hierarchy.

Proof. By definition $\Phi_{\mathcal{E}} = (\mathcal{E} \otimes \mathbb{1}_n) |\Omega_n\rangle\langle\Omega_n|$. But $|\Omega_n\rangle$ is a stabiliser state, so $\mathcal{R}(\Phi_{\mathcal{E}})$ can be no larger than $\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n) |\phi_*\rangle\langle\phi_*|) = \mathcal{C}(\mathcal{E})$, where $|\phi_*\rangle$ is the stabiliser state that achieves the capacity,

and so:

$$\mathcal{R}(\Phi_{\mathcal{E}}) \leq \mathcal{C}(\mathcal{E}). \tag{48}$$

Now suppose $\mathcal{E} = (1 + p)\Lambda_+ - p\Lambda_-$ is the optimal decomposition of $\mathcal{E}$ into CPTP stabiliser-preserving maps, $\Lambda_\pm \in \mathrm{SP}_{n,n}$, so that $\mathcal{R}_*(\mathcal{E}) = 1 + 2p$. Then for any *input* stabiliser state $\sigma \in \mathrm{STAB}_{2n}$, we can write down a valid stabiliser decomposition of the *output* state:

$$(\mathcal{E} \otimes \mathbb{1}_n)\sigma = (1 + p)(\Lambda_+ \otimes \mathbb{1}_n)\sigma - p(\Lambda_- \otimes \mathbb{1}_n)\sigma. \tag{49}$$

In particular this is true for the stabiliser state $\sigma_* = |\phi_*\rangle\langle\phi_*|$ that is optimal with respect to the capacity. But equation (49) could be a non-optimal decomposition, so its ℓ_1-norm $1 + 2p$ is at least as large as $\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)\sigma_*)$. So:

$$\mathcal{C}(\mathcal{E}) = \mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)\sigma_*) \leq 1 + 2p = \mathcal{R}_*(\mathcal{E}), \tag{50}$$

completing the proof of the first statement. Having done so, to prove the second statement it suffices to show that $\mathcal{R}(\Phi_{\mathcal{U}}) = \mathcal{R}_*(\mathcal{U})$.

For any n-qubit gate U from the third level of the Clifford hierarchy, corresponding to the channel $\mathcal{U}$, deterministic state injection is possible [7, 53]. That is, given a Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \otimes \mathcal{H}_D$, where each subspace is comprised of n qubits, there exists a completely-stabiliser-preserving circuit Λ such that, for any $2n$-qubit input state ρ:

$$\mathrm{Tr}_{BC}\left[\Lambda(\Phi_{\mathcal{U}}^{AB} \otimes \rho^{CD})\right] = \mathcal{U} \otimes \mathbb{1}_n(\rho^{AD}) \tag{51}$$

Where $\Phi_{\mathcal{U}}$ is the Choi state for the channel $\mathcal{U}$. The circuit Λ is comprised of a complete Bell measurement on BC, followed by a Clifford correction on subspace A conditioned on the outcome of the Bell measurement. It can be represented by Kraus operators:

$$K_j = (C_j^A \otimes \mathbb{1}_{3n}^{BCD})M_j, \tag{52}$$

where $M_j = \mathbb{1}^A \otimes |\Phi_j\rangle\langle\Phi_j|^{BC} \otimes \mathbb{1}^D$ are the Kraus operators corresponding to elements of the Bell basis $|\Phi_j\rangle$, and C_j is some unitary Clifford correction.

Now consider an optimal decomposition of the Choi state:

$$\Phi_{\mathcal{U}} = (1 + p)\rho_+ - p\rho_+, \quad \text{s.t.} \quad \mathcal{R}(\Phi_{\mathcal{U}}) = 1 + 2p. \tag{53}$$

We now show that by substitution into equation (51) we can obtain a decomposition of the channel that satisfies the trace-preservation condition required for channel robustness. We have:

$$\Phi_{\mathcal{U}}^{AD} = \mathcal{U} \otimes \mathbb{1}_n(|\Omega\rangle\langle\Omega|^{AD}) = \mathrm{Tr}_{BC}\left[\Lambda(\Phi_{\mathcal{U}}^{AB} \otimes |\Omega\rangle\langle\Omega|^{CD})\right] \tag{54}$$

$$= (1 + p)\widetilde{\rho}_+^{AD} - p\widetilde{\rho}_-^{AD}, \tag{55}$$

where $\widetilde{\rho}_\pm^{AD} = \mathrm{Tr}_{BC}[\Lambda(\rho_\pm^{AB} \otimes |\Omega\rangle\langle\Omega|^{CD})] \in \mathrm{STAB}_{2n}$, since $\Lambda \in \mathrm{SP}_{n,n}$. If we can show that $\mathrm{Tr}_A(\widetilde{\rho}_\pm) = \mathbb{1}_n/2^n$, then we have satisfied the required condition. First, note that $\mathrm{Tr}_A(\widetilde{\rho}_\pm)$ is independent of the Clifford corrections C_j, since the partial trace depends only on the outcome probabilities of the Bell measurement, $p_j = \mathrm{Tr}\left[M_j(\rho_\pm^{AB} \otimes |\Omega\rangle\langle\Omega|^{CD})M_j^\dagger\right]$. Therefore $\mathrm{Tr}_A(\widetilde{\rho}_\pm) = \mathrm{Tr}_{ABC}(\rho_\pm')$, where

$\rho'_\pm = \sum_j M_j(\rho_\pm^{AB} \otimes |\Omega\rangle\langle\Omega|^{CD})M_j^\dagger$ is the state following the Bell measurement. We then have:

$$\text{Tr}_A(\widetilde{\rho}_\pm) = \sum_j \text{Tr}_{ABC}\left[M_j\left(\rho_\pm^{AB} \otimes |\Omega\rangle\langle\Omega|^{CD}\right)M_j^\dagger\right] \tag{56}$$

$$= \text{Tr}_{ABC}\left[\sum_j M_j^\dagger M_j\left(\rho_\pm^{AB} \otimes |\Omega\rangle\langle\Omega|^{CD}\right)\right] \tag{57}$$

$$= \text{Tr}_{ABC}\left[\rho_\pm^{AB} \otimes |\Omega\rangle\langle\Omega|^{CD}\right] \tag{58}$$

$$= \text{Tr}_C\left[|\Omega\rangle\langle\Omega|^{CD}\right] = \frac{\mathbb{1}_n}{2^n}. \tag{59}$$

In going to the second line, we used the fact that the partial trace over BC is cyclic with respect to operators that act non-trivially only on $\mathcal{H}_B \otimes \mathcal{H}_C$. In going from the second to the third line, we used the fact that $\{M_j\}$ is a complete set of Kraus operators, so $\sum_j M_j^\dagger M_j = \mathbb{1}$. We have shown that the decomposition (55) satisfies the trace preservation criterion. Since the decomposition may not be optimal, we have that $\mathcal{R}_*(\mathcal{U}) \le 1 + 2p = \mathcal{R}(\Phi_\mathcal{U})$. But from the proof of the first statement $\mathcal{R}(\Phi_U) \le \mathcal{C}(\mathcal{U}) \le \mathcal{R}_*(\mathcal{U})$, so it must be that equality holds. $\qquad\square$

We note that the result that $\mathcal{R}(\Phi_\mathcal{U}) = \mathcal{R}_*(\mathcal{U})$ for third-level gates carries over to the case of decompositions of $\mathcal{U}(|+\rangle\langle+|^{\otimes n})$ for diagonal third-level gates. That is, there always exists a decomposition satisfying the constraints of equation (24) that is optimal with respect to $\mathcal{R}(\Phi_\mathcal{U}) = \mathcal{R}(\mathcal{U}(|+\rangle\langle+|^{\otimes n}))$. This can be seen by following the argument of Theorem 3.2, but replacing the full $4n$-qubit teleportation circuit with a $2n$-qubit state injection circuit (Figure 2).

7. CLASSICAL SIMULATION ALGORITHMS

Here we propose two classical simulation algorithms. The channel robustness $\mathcal{R}_*$ relates to the runtime of our first simulator, which we call the *static* simulator. The magic capacity $\mathcal{C}$ relates to the runtime of the second simulator, called the *dynamic* simulator. In both cases, we consider a circuit composed from a sequence of channels with $\{\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_L\}$ acting on an initial stabiliser state, which we take to be $|0^n\rangle$. The circuit ends with some final state $\rho = \mathcal{E}_L \ldots \circ \mathcal{E}_2 \circ \mathcal{E}_1(|0^n\rangle\langle 0^n|)$ and measurement of some Pauli observable Z. We assume that each channel acts non-trivially on a bounded number of qubits (e.g. 2 or 3) so we can evaluate the relevant monotones. Our goal is to estimate the expectation value $\text{Tr}[Z\rho]$ to within additive error. In the language of Ref. [54] our simulators will be poly-boxes.

Both our algorithms are inspired by previous methods that collect a large number of Monte Carlo samples that scales quadratically with the negativity of some quasiprobability distribution [3, 33, 38]. The static Monte Carlo simulator uses a precomputed, and therefore static, quasiprobability distribution. The dynamic Monte Carlo simulator recomputes optimal quasiprobability distributions at each step, which can lead to fewer samples required but with a higher runtime per sample. As such, there are subtle trade-offs in the runtime complexities.

Both our algorithms use that completely stabiliser preserving operations $\text{SP}_{n,n}$ acting on a stabiliser state can be classically efficiently simulated. This follows from the fact that given the Choi state $\Phi_\mathcal{E}$ for an n-qubit channel, the channel may be implemented by performing a Bell measurement on $\Phi_\mathcal{E} \otimes \sigma$, postselecting on the Ω outcome to obtain $(\mathbb{1}_n \otimes |\Omega\rangle\langle\Omega|)\Phi_\mathcal{E} \otimes \sigma(\mathbb{1}_n \otimes |\Omega\rangle\langle\Omega|)$ and then tracing out the last $2n$ qubits [53]. This can be simulated using Gottesman-Knill when σ is a stabiliser state and $\mathcal{E} \in \text{SP}_{n,n}$. Curiously, it is unclear whether $\text{SP}_{n,n}$ can be physically realised using Clifford unitaries and Pauli measurements but without the use of postselection.

A. Static Monte Carlo

In Algorithm 1 we give pseudocode for our first simulator, which we call static Monte Carlo, and which can be viewed as a generalisation of the algorithm of Bennink et al. [3], differing in two important ways. First, whereas their algorithm involved an optimisation over the set CPR, the channel robustness is optimised with respect to $\mathrm{SP}_{m,m}$, so that all completely stabiliser-preserving maps are represented non-negatively. Second, while the positive and negative parts of the decomposition are each CPTP, the subroutine in step (c) of the algorithm involves updating the state with maps that may not be trace-preserving. This is necessary to ensure that the updated state after step (c)iii. remains pure. We explain below how this subroutine is carried out and show that it does not increase the sample complexity.

We assume a pre-computation stage in which for each circuit element we determine an optimal decomposition as per the definition (5), so that:

$$\mathcal{E}_j = (1 + p_j)\mathcal{E}_{j,0} - p_j\mathcal{E}_{j,1}, \quad \text{where} \quad \mathcal{R}_*(\mathcal{E}_j) = 1 + 2p_j \tag{60}$$

where $\mathcal{E}_{j,k} \in \mathrm{SP}_{m,m}$. The runtime of this pre-computation stage is bounded since each circuit element acts non-trivially on only m qubits. Once these decompositions have been calculated, one can define an overall quasiprobability distribution:

$$q_{\vec{k}} = \prod_{j:k_j=0} (1 + p_j) \prod_{j:k_j=1} (-p_j), \tag{61}$$

where $\vec{k} \in \mathbb{Z}_2^L$ is a vector representing a choice of either $\mathcal{E}_{j,0}$ or $\mathcal{E}_{j,1}$ at each circuit element $\mathcal{E}_j$. We can renormalise this to obtain a product probability distribution:

$$p_{\vec{k}} = \prod_{j:k_j=0} \frac{(1 + p_j)}{\mathcal{R}_*(\mathcal{E}_j)} \prod_{j:k_j=1} \frac{(p_j)}{\mathcal{R}_*(\mathcal{E}_j)}. \tag{62}$$

Hence we can write the output of the circuit as follows:

$$\mathcal{E}(|\phi_0\rangle\langle\phi_0|) = R \sum_{\vec{k}} p_{\vec{k}}\lambda_{\vec{k}}\Lambda_{\vec{k}}(|\phi_0\rangle\langle\phi_0|) \tag{63}$$

where each $\Lambda_{\vec{k}} = \mathcal{E}_{L,k_L} \circ \ldots \circ \mathcal{E}_{1,k_1}$ gives a trajectory of $\mathrm{SP}_{m,m}$ channels through the circuit, $\lambda_{\vec{k}} = \mathrm{sign}(q_{\vec{k}})$, and $R = \prod_{j=1}^L \mathcal{R}_*(\mathcal{E}_j)$. The expectation value of an observable Z at the end of the circuit is given by:

$$\mathrm{Tr}[Z\mathcal{E}(|\phi_0\rangle\langle\phi_0|)] = R \sum_{\vec{k}} p_{\vec{k}}\lambda_{\vec{k}} \mathrm{Tr}\left[Z\Lambda_{\vec{k}}(|\phi_0\rangle\langle\phi_0|)\right] \tag{64}$$

This decomposition is a quasi-probability distribution with ℓ_1-norm

$$||\vec{q}||_1 = \sum_{\vec{k}} |p_{\vec{k}}\lambda_{\vec{k}}R| = \sum_{\vec{k}} |p_{\vec{k}}|R = \prod_j \mathcal{R}_*(\mathcal{E}_j). \tag{65}$$

By sampling from the stabiliser-preserving trajectories $\Lambda_{\vec{k}}$ with probability distribution $\{p_{\vec{k}}\}$, and calculating $\lambda_{\vec{k}}R \, \mathrm{Tr}\left[Z\Lambda_{\vec{k}}(|\phi_0\rangle\langle\phi_0|)\right]$, we have an unbiased estimator for $\mathrm{Tr}[Z\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]$. Crucially, (62) gives a product distribution, so this first sampling step is efficient. The variance is increased by $R > 1$, but by standard arguments [3, 33, 38] the Hoeffding inequalities can be used to show that any constant error $\delta > 0$ in the mean estimate can be achieved with arbitrary success probability

$(1 - \epsilon)$ by repeating the sampling procedure N times, where:

$$N = \frac{2}{\delta} ||\vec{q}||_1^2 \ln\left(\frac{2}{\epsilon}\right).$$ (66)

Algorithm 1 A classical simulator with sample complexity $\prod_{j=1}^{L} \mathcal{R}_*(\mathcal{E}_j)^2$.

Input: Circuit description $\{\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_L\}$, where each $\mathcal{E}_j$ has a known optimal decomposition with channel robustness $\mathcal{R}_*(\mathcal{E}_j)$ as per equations (60) and (67), input stabiliser state $|\phi_0\rangle$, Pauli observable Z, number of samples M.

Output: Estimate of the Pauli expectation value $\langle Z \rangle = \text{Tr}[Z\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]$

1. Set $i \leftarrow 1$, $T \leftarrow 0$ and $R \leftarrow \prod_{j=1}^{L} \mathcal{R}_*(\mathcal{E}_j)$;

2. For $i = 1$ to M:

    (a) Sample vector $\vec{k}_i$ according to the distribution $\{p_{\vec{k}}\}$;

    (b) Set the input state of the circuit to be the stabiliser state $|\phi_0\rangle$;

    (c) For $j = 1$ to L:

        i. Calculate the distribution:

        $$\left\{ p_l^{(j)} = \text{Tr}\left[\mathcal{T}_{l,j,k_j}(|\phi_{j-1}\rangle\langle\phi_{j-1}|) \right] \right\};$$

        ii. Sample l with probability $p_l^{(j)}$;

        iii. Set $|\phi_j\rangle\langle\phi_j| = \mathcal{T}_{l,j,k_j}(|\phi_{j-1}\rangle\langle\phi_{j-1}|)/p_l^{(j)}$.

    (d) Calculate $\langle P \rangle_i = \text{Tr}[P|\phi_L\rangle\langle\phi_L|]$;

    (e) Set $T_i = \text{sign}\left(q_{\vec{k}_i}\right) R \langle P \rangle_i$;

    (f) $T \leftarrow T + T_i$.

Return: T/M

The simulator proceeds by tracking the evolution of a pure stabiliser state through the sampled trajectory $\Lambda_{\vec{k}}$, as described in Algorithm 1. In practice, while each $\mathcal{E}_{j,k_j}$ is constrained to be CPTP, the output of the linear program will be a decomposition of $\mathcal{E}_{j,k_j}$ into maps that are stabiliser-preserving but not necessarily trace-preserving, corresponding to Choi states that are proportional to pure stabiliser projectors:

$$\mathcal{E}_{j,k_j} = \sum_l \mathcal{T}_{l,j,k_j} \quad \longleftrightarrow \quad \Phi_{\mathcal{E}_{j,k_j}} = \sum_l |q_{l,j,k_j}| \, |\psi_{l,j,k_j}\rangle\langle\psi_{l,j,k_j}|.$$ (67)

Nevertheless, because $\mathcal{E}_{j,k_j}$ is a CPTP channel, it can be simulated by sampling from a proper probability distribution as defined in step (c)i. of Algorithm 1. We stress that while $\Phi_{\mathcal{E}_{j,k_j}}$ is subject to the constraint ensuring that each $\mathcal{E}_{j,k_j}$ is trace-preserving, this need not apply to the individual terms $|\psi_{l,j,k_j}\rangle\langle\psi_{l,j,k_j}|$ in its decomposition. Indeed, any decomposition of a non-unital channel cannot be represented as a linear combination of unitary Clifford operations alone. While Clifford gates are represented in the distribution by maximally entangled states, product states correspond to non-trace-preserving maps involving projections. For example, $\Phi_{\mathcal{T}} = |01\rangle\langle01|$ corresponds to the single Kraus operator $\mathcal{T} = |0\rangle\langle1|$ which can be seen as a Z-measurement post-selected on the "$|1\rangle$"

outcome, followed by an X gate. Under these well-defined stabiliser operations, the state update in step (c)iii. can be carried out efficiently as per the Gottesman-Knill theorem [1, 2]. More generally, the state update corresponding to any pure Choi state $|\psi_{l,j,k_j}\rangle$ can be modelled as a post-selected Bell measurement [53]:

$$|\phi_j\rangle^A \otimes |\Omega_n\rangle^{B|C} \propto (\mathbb{1}_n^A \otimes |\Omega_n\rangle\langle\Omega_n|^{B|C}) \, |\psi_{l,j,k_j}\rangle^{AB} \otimes |\phi_{j-1}\rangle^C. \tag{68}$$

Note that the new state $|\phi_j\rangle$ is guaranteed to be pure, since the projection of the system BC onto a Bell state removes any correlation across the partition $A|BC$. We emphasise that the probability for carrying out this update depends not just on the known prefactors $|q_{l,j,k_j}|$, but on $\text{Tr}\big[\mathcal{T}_{l,j,k_j}(|\phi_{j-1}\rangle\langle\phi_{j-1}|)\big]$, which are input state dependent. Computing this trace amounts to evaluating the overlap between $(\mathbb{1}_n \otimes |\Omega_n\rangle\langle\Omega_n|)$ and some stabiliser state, as per equation (68), which can be done efficiently using the stabiliser tabulex method [1, 2]. However, since $|\phi_{j-1}\rangle$ will have been chosen randomly in the previous step of the algorithm, these traces cannot be calculated ahead of time. This means that our algorithm involves additional per-sample runtime overhead in order to calculate $\{p_l^{(j)}\}$. However, there are a finite and bounded number of such calculations since each circuit element acts on a small number of qubits. The key point is that the constraint on $\mathcal{E}_{j,k_j}$ ensures that, despite being composed of non-trace-preserving elements, the distribution over the outcomes $\{p_l^{(j)}\}$ for any given state forms a proper probability distribution. Consequently, the intermediate sampling in step (c)i. does not increase the variance, so the sample complexity of the simulator depends only on $\prod_j \mathcal{R}_*(\mathcal{E}_j)^2$.

Suppose we want to compare the runtime of our simulator for a particular circuit with that of the Bennink et al. algorithm [3]. Their decompositions are in terms of CPR, the set of Cliffords and Pauli reset channels, and an associated cost function is

$$\mathcal{R}_{\text{CPR}}(\mathcal{E}) = \min_{\Lambda_j \in \text{CPR}} \left\{ ||p||_1 : \sum p_j \Lambda_j = \mathcal{E} \right\}, \tag{69}$$

The sample complexity for simulating a given circuit element $\mathcal{E}_j$ is proportional to $\mathcal{R}_{\text{CPR}}(\mathcal{E}_j)^2$. Since $\text{CPR} \subseteq \text{SP}_{n,n}$, it must be the case that $\mathcal{R}_* \leq \mathcal{R}_{\text{CPR}}$, potentially leading to lower simulation sample complexity if there exist channels with $\mathcal{R}_* < \mathcal{R}_{\text{CPR}}$. We give here a simple toy example demonstrating a significant advantage to our static simulator.

Consider the single-qubit CPTP map Λ_H defined by a Z-measurement followed by a Hadamard gate conditioned on the "-1" outcome. This has Kraus representation:

$$K_1 = |0\rangle\langle 0|, \quad K_2 = |-\rangle\langle 1|. \tag{70}$$

This is clearly a completely stabiliser-preserving map, so has channel robustness $\mathcal{R}_*(\Lambda_H) = 1$. For a single qubit, CPR consists of the 24 Clifford gates, and 6 Pauli reset channels. Using this set, we calculate $\mathcal{R}_{\text{CPR}}(\Lambda_H) = 2$. Since $\Lambda_H \in \text{SP}_{1,1}$, this confirms that CPR is a strict subset of the completely stabiliser-preserving channels, and indicates that $\mathcal{R}_{\text{CPR}}$ is not a monotone under stabiliser operations. We also note that the calculated value is larger than the robustness of magic for any single-qubit state, despite Λ_H being a stabiliser operation. For a circuit containing M uses of the channel Λ_H, the samples required for a CPR simulator would be proportional to $\mathcal{R}_{\text{CPR}}(\Lambda_H)^{2M} = 4^M$. But for our simulator, Λ_H can be simulated efficiently, as $\mathcal{R}_*(\Lambda_H)^{2M} = 1$.

While the above example is quite artificial, a reduction in sample complexity is also achieved for channels where $\mathcal{R}_*(\mathcal{E}) > 1$, but is strictly smaller than $\mathcal{R}_{\text{CPR}}(\mathcal{E})$. Given a circuit decomposed as $\mathcal{E} = \mathcal{E}_L \circ \ldots \circ \mathcal{E}_2 \circ \mathcal{E}_1$, the sample complexity for the CPR simulator would be proportional to $\prod_j^L \mathcal{R}_{\text{CPR}}(\mathcal{E}_j)^2$. It is always the case that $\mathcal{R}_*(\mathcal{E}_j) \leq \mathcal{R}_{\text{CPR}}(\mathcal{E}_j)$, so the sample complexity for our simulator will never be greater. Suppose we find that there are M circuit elements such that $\mathcal{R}_*(\mathcal{E}_j)/\mathcal{R}_{\text{CPR}}(\mathcal{E}_j) \leq k$ for some constant $0 < k < 1$. Then we would find that using our simulator gives a reduction in sample complexity by a factor of k^{2M}. While our simulator sometimes incurs

a modest increase in the runtime per sample, this must be weighed against a reduction in sample complexity that is exponential in the number of circuit elements where $\mathcal{R}_*(\mathcal{E}_j) < \mathcal{R}_{\mathrm{CPR}}(\mathcal{E}_j)$. The obvious next question is whether there are any natural non-trivial examples where this happens. We show in Section 8 that gate sequences subject to amplitude-damping noise provide one such case.

We also note that calculation of optimal CPR decompositions is only tractable for one- and two-qubit circuit elements, as the three-qubit case already involves a linear program with nearly 93 million variables [3]. For the most general quantum channels, we encounter a similar problem, as for three-qubit circuit elements, we in principle need to optimise over six-qubit stabiliser states. However, in Appendix E we show that for diagonal channels the problem can be greatly simplified, and the problem becomes tractable for operations on up to five qubits. This allows our algorithm to take advantage of the submultiplicativity of channel robustness; for example for diagonal channels where $\mathcal{R}(\mathcal{E}^{\otimes n}) < \mathcal{R}(\mathcal{E})^n$, it is advantageous to compose n single-qubit circuit elements together as a single n-qubit circuit element, before running the linear program. We will see in Section 8 that this strategy is useful for single-qubit Z-rotations.

B. Dynamic Monte Carlo

In the previous simulator, all convex optimisations are calculated in advance. However, we have found examples of channels where $\mathcal{C}(\mathcal{E}) < \mathcal{R}_*(\mathcal{E})$. For such channels, and for any stabiliser state ρ, the robustness of the output state $\mathcal{E}(\rho)$ will always be less than the ℓ_1-norm of the decomposition of $\mathcal{E}$ into stabiliser-preserving CPTP channels. Our next simulator takes advantage of this, and we present pseudocode in Algorithm 2.

As in the previous simulator, we can represent the trajectory through the circuit by a vector $\vec{k}$, such that the output of the true quantum circuit would be $\rho = \sum_{\vec{k}} q_{\vec{k}} \sigma_{\vec{k}}$. The major difference is that $q_{\vec{k}}$ cannot be written in the form of equation (62) with p_j independent of $\vec{k}$, as the quasiprobabilities for each intermediate decomposition will depend on the stabiliser state sampled in the previous step. Consider what happens for the jth circuit element. After step (d)ii. of the algorithm, we have some stabiliser state $\sigma_{\vec{k}_j}$, where $\vec{k}_j$ labels the trajectory up to the jth element. Here we do not assume $\vec{k}_j$ is a binary vector; instead the elements of the vector label each pure stabiliser state. After steps (d)iii. and iv. we have a non-stabiliser state $\rho_{\vec{k}_j} = (\mathcal{E}_j \otimes \mathbb{1})(\sigma_{\vec{k}_j})$, decomposed as:

$$\rho_{\vec{k}_j} = \sum_{\vec{k}_{j+1}} q_{\vec{k}_{j+1}} \sigma_{\vec{k}_{j+1}}. \tag{71}$$

Here the summation is over all $(j+1)$-step trajectories consistent with the previous j-step trajectory labelled by $\vec{k}_j$. In steps (d)v. and vi. we will choose the stabiliser state $\sigma_{\vec{k}_{j+1}}$ with probability $|q_{\vec{k}_{j+1}}|/\mathcal{R}(\rho_{\vec{k}_j})$ and the variable R picks up a factor $\lambda_{\vec{k}_{j+1}} \mathcal{R}(\rho_{\vec{k}_j})$, where $\lambda_{\vec{k}_{j+1}}$ is the sign of the corresponding quasiprobability.

The final state after the full sequence of quantum operations may be written:

$$\rho = \sum_{\vec{k}} p_{\vec{k}} R_{\vec{k}} \sigma_{\vec{k}}, \quad \text{where} \quad p_{\vec{k}} = \prod_{j=0}^{L-1} \frac{|q_{\vec{k}_{j+1}}|}{\mathcal{R}(\rho_{\vec{k}_j})}, \quad R_{\vec{k}} = \prod_{j=0}^{L-1} \mathrm{sign}\left(q_{\vec{k}_{j+1}}\right) \mathcal{R}(\rho_{\vec{k}_j}). \tag{72}$$

The true expectation value for the observable Z is given by:

$$\mathrm{Tr}[Z\rho] = \sum_{\vec{k}} p_{\vec{k}} R_{\vec{k}} E_{\vec{k}}, \quad \text{where} \quad E_{\vec{k}} = \mathrm{Tr}\left[Z\sigma_{\vec{k}}\right]. \tag{73}$$

Algorithm 2 A classical simulator with sample complexity per step upper-bounded by $\mathcal{C}^2$.

Input: A sequence of L quantum channels $\mathcal{E}_j$ and number of samples N.

Output: An estimate of an expectation value.

1. Set $i \leftarrow 0$ and $T \leftarrow 0$;

2. For $i \leq N$ do:

   (a) $i \leftarrow i + 1$;

   (b) Set $|\phi_0\rangle = |0\rangle^{\otimes n}$;

   (c) $R \leftarrow 1$;

   (d) For $1 \leq j \leq L$ do

      i. The channel $\mathcal{E}_j$ acts non-trivially on only m qubits. Partition the qubits into three sets $A|B|C$ where A is the set acted on by $\mathcal{E}_j$, B is a set of any other m qubits; and C comprises the remaining qubits;

      ii. Find a Clifford $U = \mathbb{1}^A \otimes U^{BC}$ that is local w.r.t $A|(B \cup C)$ such that $U |\phi_{j-1}\rangle = \left|\phi_{j-1}^{AB}\right\rangle \otimes \left|\phi_{j-1}^C\right\rangle$. This uses the efficient algorithm of Ref. [52].

      iii. Find $2m$-qubit density matrix $\rho_j^{AB} = \mathcal{E}_j \otimes \mathbb{1}\left(\left|\phi_{j-1}^{AB}\right\rangle\left\langle\phi_{j-1}^{AB}\right|\right)$;

      iv. Solve convex optimisation to find $\mathcal{R}(\rho_j^{AB})$ and use optimal decomposition to build a quasiprobability distribution ;

      v. Sample from the renormalised quasiprobability distribution to choose a stabiliser state $\left|\phi_j'^{AB}\right\rangle$, and then set $|\phi_j\rangle = U^\dagger \left|\phi_j'^{AB}\right\rangle \otimes \left|\phi_{j-1}^C\right\rangle$;

      vi. Replace $R \leftarrow R \times \mathcal{R}(\rho_j^{AB}) \times \lambda$ where $\lambda = \pm 1$ and denotes the phase of the sampled quasiprobability.

      vii. increment $j \leftarrow j + 1$ and loop;

   (e) Evaluate $E = \langle\phi_L| Z |\phi_L\rangle$

   (f) $T \leftarrow T + (R \times E)$.

Return: T/N.

A key difference with the previous simulator is that we never explicitly calculate the full distribution $p_{\vec{k}}$. Nevertheless, each time the simulator samples, it produces output $R_{\vec{k}}E_{\vec{k}}$ with probability $p_{\vec{k}}$. These probabilities exactly match the weightings in the above equation, so the simulator is an unbiased estimator. The number of samples required can be again derived using the Hoeffding inequalities, which depend on the maximum possible values of the output of each sample. Each output is bounded by $|R_{\vec{k}}E_{\vec{k}}| \leq |R_{\vec{k}}| \leq \prod_j \mathcal{R}(\rho_{\vec{k}_j}) \leq \prod_j \mathcal{C}(\mathcal{E}_j)$. Therefore, the sample complexity is upper-bounded by order $\prod_j \mathcal{C}(\mathcal{E}_j)^2$.

Notice that for every sample, L convex optimisations are performed, as well as L steps involving the algorithm of Fattal et al. [52], which has runtime polynomial in the total number of qubits. If $\mathcal{C}(\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$ then we would simply not use this method so that the only convex optimisations are in the preprocessing. However, $\mathcal{C}(\mathcal{E})$ now determines the sample complexity, so if $\mathcal{C}(\mathcal{E}) \ll \mathcal{R}_*(\mathcal{E})$ then the dynamic simulator may run much faster than the static simulator; here we have a trade-off of increase in per-sample runtime, versus a possibly exponential reduction in sample complexity. Indeed, since the sample complexity is typically the bottleneck, this approach would lead to significant improvements for some quantum channels. In the following section we investigate

which types of channels may lead to an advantage.

8.   NUMERICAL RESULTS

The numerical results in this section have been produced using code and data files available from the public repository detailed in Ref. [44].

A.   Single-qubit rotation with amplitude damping

Consider the setting discussed in Section 7, where a many-qubit circuit evolution $\mathcal{E}$ is decomposed as a series of few-qubit circuit elements $\mathcal{E} = \mathcal{E}_L \ldots \circ \mathcal{E}_2 \circ \mathcal{E}_1$. Many implementations of quantum algorithms can be expected to involve single-qubit rotations about some Pauli axis. In near-term devices, the circuit will be subject to noise. Consider a simple model of a noisy computation where a noise channel Λ acts between each unitary gate $\mathcal{U}_j$, so the overall channel representing the circuit would be:

$$\mathcal{E} = \Lambda \circ \mathcal{U}_L \circ \ldots \circ \Lambda \circ \mathcal{U}_2 \circ \Lambda \circ \mathcal{U}_1. \tag{74}$$

Here we study the simulation cost for a single step in such a computation, comprised of a single-qubit rotation and a noise channel. Note that for intermediate steps in a circuit decomposition such as (74), we have a choice of ordering. We can take the circuit elements to be either $\Lambda \circ \mathcal{U}_j$ or $\mathcal{U}_j \circ \Lambda$. These choices are equivalent in terms of the output of the simulation, but could lead to different sample complexity depending on the cost function used. We studied circuit elements made up of a single-qubit Pauli X-rotation $U(\theta) = \exp(iX\theta)$ composed with an amplitude damping channel Λ_p with noise parameter p, defined by Kraus operators:

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix} \tag{75}$$

We calculated channel robustness $\mathcal{R}_*$, magic capacity $\mathcal{C}$, and the Bennink et al. [3] cost function $\mathcal{R}_{\mathrm{CPR}}$, for both $\Lambda_p \circ \mathcal{U}(\theta)$ and $\mathcal{U}(\theta) \circ \Lambda_p$, and for a range of values of p and θ. For noise $p = 0.1$ we see that when the noise channel follows the gate, there is no difference between the three quantities (Figure 3(i)). However, if the noise channel acts before the unitary, both our monotones show a reduced value, whereas $\mathcal{R}_{\mathrm{CPR}}$ increases. This suggests that for this noise model, the better strategy with respect to sample complexity would be to choose the ordering $\mathcal{U}(\theta) \circ \Lambda_p$, and use one of our simulators (subject to the caveats mentioned in the previous section). In Figure 3(ii) we show how different levels of noise affect the channel robustness. We do not plot capacity since we find that $\mathcal{R}_*(\mathcal{E}) = \mathcal{C}(\mathcal{E})$ for this class of operation, up to solver precision. We also compare channel robustness with the Choi state robustness (Figure 3(iii)). We find that $\mathcal{R}(\Phi_{\mathcal{E}}) = \mathcal{R}_*(\mathcal{E})$ for θ up to approximately $\pi/16$, but $\mathcal{R}(\Phi_{\mathcal{E}}) < \mathcal{R}_*(\mathcal{E})$ for larger angles.

B.   Multiqubit phase gates

Recall that from Theorem 3.2 we know $\mathcal{R}(\Phi_{\mathcal{E}}) \leq \mathcal{C}(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})$. For the particular class of channel studied above, we saw numerically that $\mathcal{C}(\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$ up to solver precision in all cases investigated, and in the absence of noise the numerical results suggested $\mathcal{R}(\Phi_{\mathcal{E}}) = \mathcal{C}(\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$. Moreover, we know from Theorem 3.2 that all measures are equal for gates from the third level of the Clifford hierarchy. Under what conditions does this equality persist for multi-qubit operations? As explained earlier, to calculate each of our quantities for n-qubit channels, in general we must solve an optimisation problem over all $2n$-qubit stabiliser states. Since this problem is only tractable for
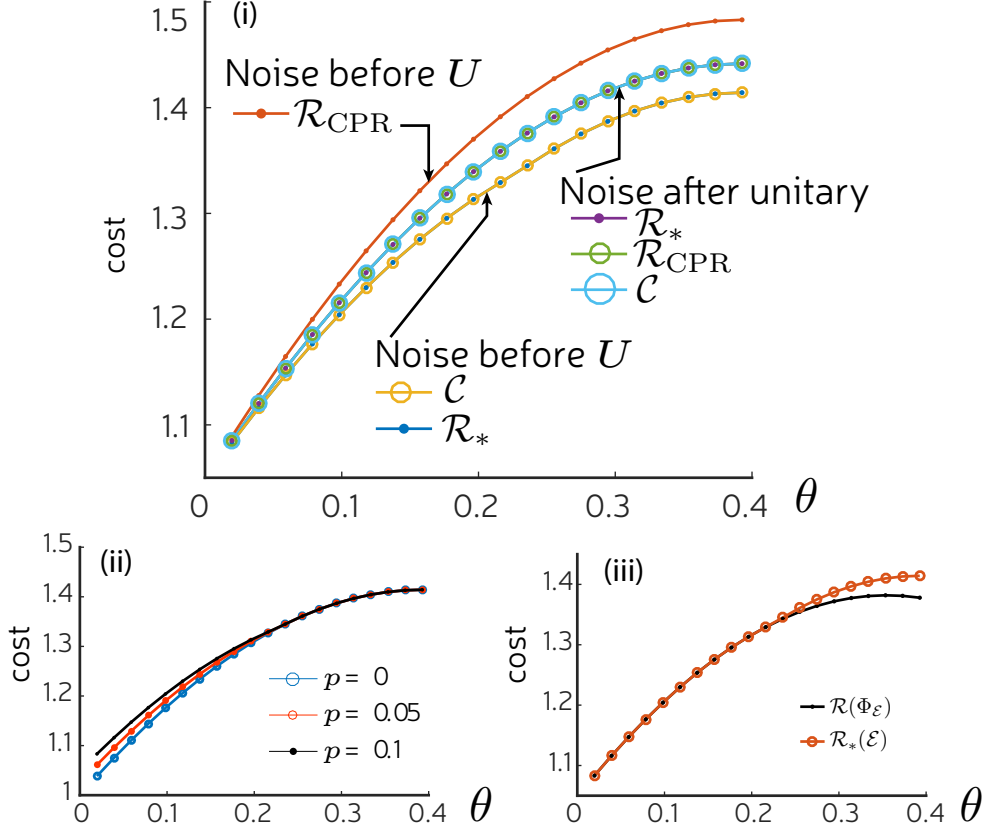
FIG. 3. (i) Comparison of $\mathcal{C}(\mathcal{E})$, $\mathcal{R}_*(\mathcal{E})$ and $\mathcal{R}_{CPR}(\mathcal{E})$, where $\mathcal{E}$ is a single-qubit X-rotation $\mathcal{U}(\theta)$ composed with an amplitude damping channel Λ_p. We consider both possible orderings: noise after unitary ($\Lambda_p \circ \mathcal{U}(\theta)$), and noise before unitary ($\mathcal{U}(\theta) \circ \Lambda_p$). (ii) $\mathcal{R}_*(\mathcal{U}(\theta) \circ \Lambda_p)$ for several values of p. (iii) Comparison of channel robustness with robustness of Choi state for $\mathcal{U}(\theta) \circ \Lambda_p$ with $p = 0.1$.

up to 5-qubit states, in practice we are limited to studying two-qubit channels, in the most general case. However, it turns out the problem can be greatly simplified for certain types of operation. In particular, Appendix E shows how the problem size can be reduced for channels diagonal in the computational basis, using a representation of stabiliser states in terms of affine spaces over binary vectors due to Dehaene and De Moor [55, 56]. MATLAB code to calculate our measures for this reduced problem is provided in the public repository given in Ref. [44]. This allows us to calculate values for diagonal operations on up to 5 qubits, which we present here.

As a special case we consider multicontrol phase gates of the form:

$$M_{t,n} = \mathrm{diag}\big(\exp\big(i\pi/2^t\big), 1, 1, \dots, 1\big), \quad t \in \mathbb{Z} \tag{76}$$

where n denotes the number of qubits. We note that by convention, controlled-phase gates typically apply the phase to the all-one state $|1^n\rangle$, where $1^n = (1, \dots, 1)^T$, but the form given above is Clifford-equivalent to the conventional version, and is more convenient for the techniques used in Appendix E. The family includes familiar gates such as CZ ($t = 0$, $n = 2$), CCZ ($t = 0$, $n = 3$), multicontrol-S ($t = 1$) and multicontrol-T ($t = 2$).

The main findings were that the inequalities are tight for the $n = 2$ and $n = 3$ cases, but that this does not persist for larger system sizes (Figure 4). The $t = 0$ case (the family of multicontrol-Z gates) turns out to be a special case (Figure 4, left panel). Here we find equality for all three quantities up to $n = 4$. For the $t = 0$, $n = 5$ case, $\mathcal{R}\big(\Phi_{M_{0,5}}\big) = \mathcal{C}(M_{0,5})$ holds, but $\mathcal{R}_*(M_{0,5})$

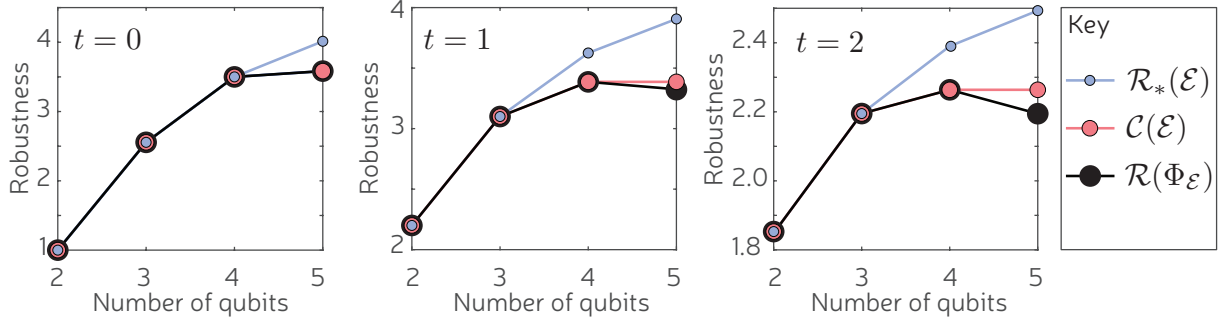is strictly greater than both. Note also that for $t = 0$, all three quantities increase with each increment in n.

FIG. 4. Comparison of quantities for multicontrol phase gates (see equation 76). Left: Multicontrol-Z gates (t=0). Middle: Multicontrol-S gates (t=1). Right: Multicontrol-T gates (t=2).
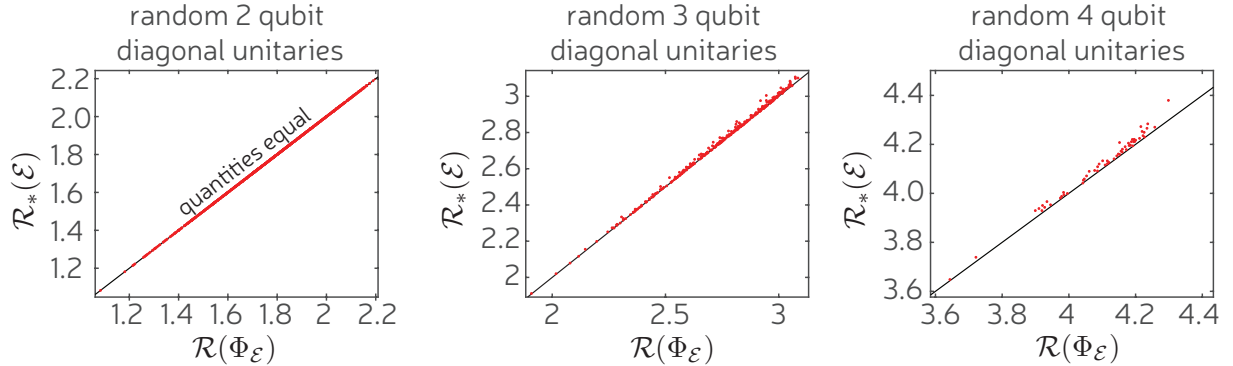

FIG. 5. Channel robustness against robustness of Choi state for random n-qubit diagonal gates, up to $n = 4$. Black line indicates $\mathcal{R}_* = \mathcal{R}(\Phi)$. Each red dot represents the data point for an individual gate. Fewer points were calculated for larger n due to the increased time to calculate each value. 1000 data points were calculated for $n = 2$, 300 for $n = 3$, and 60 for $n = 4$.

The families of gates with $t > 0$ follow a pattern qualitatively similar to each other. The results for the $t = 1$ (multicontrol-S) and $t = 2$ (multicontrol-T) cases are shown in the middle and right panels of Figure 4. For $n = 4$, $t > 0$, the same situation holds as for $n = 5$, $t = 0$, as we find $\mathcal{R}\left(\Phi_{M_{t,4}}\right) = \mathcal{C}(M_{t,4}) < \mathcal{R}_*(M_{t,4})$. At $n = 5$, all three quantities separate. In contrast with the multicontrol-Z, we see that $\mathcal{R}\left(\Phi_{M_{t,n}}\right)$ decreases as we go from four to five qubits, while $\mathcal{C}(M_{t,n})$ levels off. We see similar behaviour for all non-zero values of t investigated numerically. Our current techniques limit us to five-qubit gates, but we have reason to believe that the capacity will remain level for $n > 5$, and we make the following conjecture, which we justify more fully in Appendix E 4.

Conjecture 8.1. *For any fixed t, the maximum increase in robustness of magic for $M_{t,n}$ is achieved at some finite number of qubits $n = K$ by acting on the state $|+\rangle^{\otimes K}$. Therefore $\mathcal{C}(M_{t,n}) = \mathcal{R}\left(M_{t,K}|+\rangle^{\otimes K}\right)$ for all $n \geq K$.*

We also numerically investigated the robustness of diagonal unitaries

$$U = \sum_x e^{i\theta_x}|x\rangle\langle x|, \tag{77}$$

with θ_x chosen uniformly at random. We were particularly interested in understanding when $\mathcal{R}(\Phi_\mathcal{E}) \leq C(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})$ is tight or loose. In Figure 5 we compare the Choi robustness with the channel robustness. For every 2-qubit gate tested we observed that $\mathcal{R}(\Phi_\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$ up to numerical precision. Whereas, for 3 and 4 qubit gates we typically saw that $\mathcal{R}(\Phi_\mathcal{E}) < \mathcal{R}_*(\mathcal{E})$, though the gap is not often large. While the difference is slight for a single gate, these quantities influence the rate of exponential scaling when considering N uses of such a unitary and will lead to a large gap for modest N.

We also compared the Choi robustness with the magic capacity but do not plot this data as it was equal for every random instance we observed. This is curious since in Figure 4 we clearly see that there do exist diagonal gates, the multicontrol phase gates, for which there is a gap between the Choi robustness and the magic capacity. While such gates exist, our random sampling of diagonal gates does not tend to provide such examples. We discuss this further in Appendix E 4.

Finally, we are also interested in the *normalised* channel robustness for single-qubit gates U, defined as $[\mathcal{R}_*(U^{\otimes n})]^{1/n}$. This allows us to quantify the per-gate savings in sample complexity that can be achieved by grouping single-qubit rotations in n-qubit blocks. In Figure 6 we present results for qubit Z-rotations $U = \exp[iZ\theta]$, up to four qubits. We find that strict submultiplicativity is observed for all values of θ, with significant reductions between the $n = 2$ and $n = 4$ cases for a wide range of angles.
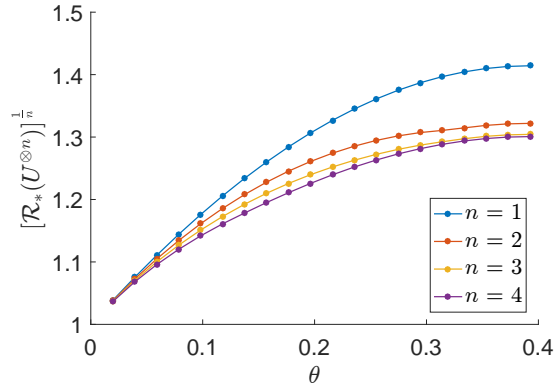

FIG. 6. Normalised channel robustness $[\mathcal{R}_*(U^{\otimes n})]^{1/n}$ plotted for Z-rotations $U(\theta) = \exp[iZ\theta]$ and for n qubits, up to $n = 4$.

9.   CONCLUSION

We have presented two new magic monotones for general quantum channels: the magic capacity $\mathcal{C}$, which quantifies the ability of a channel to generate magic, and the channel robustness $\mathcal{R}_*$, which is related to finding the minimal quasiprobability decomposition of a channel into stabiliser-preserving CPTP maps. Each of these monotones is directly related to the sample complexity for an associated Monte Carlo-type classical simulation algorithm. We found that for certain quantum channels, our static simulator would lead to a exponentially better sample complexity as compared to that for the algorithm due to Bennink et al. [3]. In particular we found a reduction in sample complexity for the case of a sequence of single-qubit rotations subject to amplitude-damping noise. Since our decompositions can be calculated for up to five qubits in the case of diagonal operations,

our static simulator is also able to take advantage of the submultiplicativity of channel robustness under tensor product and composition.

For some channels, further improvements in sample complexity are possible using a different simulator that is related to the capacity. That is, we found the capacity can be strictly less than channel robustness for certain multi-qubit entangling gates. This simulator has to introduce on-the-fly convex optimisation, however, so each sample will be more difficult to obtain.

For simulation of realistic quantum devices involving many qubits, one would need to decompose the circuit into a sequence of operations on smaller number of qubits, as described in Section 7, in order that the associated optimisation problems are tractable. It is a non-trivial problem to decide what is the optimal way to block together the few-qubit operations making up a given many-qubit circuit: we saw in Section 8 that the ordering of operations can make a difference to the sample complexity. We leave this problem for a future work.

Since our monotones are submultiplicative under tensor product and compositions, it is generally preferable to combine circuit elements where possible. In practice, to calculate our monotones, circuit elements can involve up to two qubits for the most general case, or at most five qubits for diagonal operations, where we can make use of the techniques described in Appendix E. In Ref. [47], Heinrich and Gross show that robustness of magic can be calculated for up to 10 copies of the resource states employed in the standard magic state model of fault-tolerant quantum computation. Their methods rely on both the permutation symmetry due to having multiple copies, and the stabiliser symmetries of the magic states considered. Another direction for future work could therefore be to investigate whether (and under what conditions) similar techniques can be applied to the channel picture to increase the number of qubits that can be involved in each circuit element.

[1] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Ph.D. thesis, Caltech (1997), arXiv:quant-ph/9705052.
[2] S. Aaronson and D. Gottesman, Physical Review A **70**, 052328 (2004), arXiv:quant-ph/0406196.
[3] R. S. Bennink, E. M. Ferragut, T. S. Humble, J. A. Laska, J. J. Nutaro, M. G. Pleszkoch, and R. C. Pooser, Physical Review A **95**, 062337 (2017), arXiv:1703.00111.
[4] M. Yoganathan, R. Jozsa, and S. Strelchuk, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **475**, 20180427 (2019), arXiv:1806.03200.
[5] J. Preskill, Quantum **2**, 79 (2018), arXiv:1801.00862.
[6] E. Knill, Nature **434**, 39 (2005), arXiv:quant-ph/0410199.
[7] S. Bravyi and A. Kitaev, Physical Review A **71**, 022316 (2005), arXiv:quant-ph/0403025.
[8] E. T. Campbell, B. M. Terhal, and C. Vuillot, Nature **549**, 172 (2017), arXiv:1612.07330.
[9] B. W. Reichardt, Quantum Information Processing **4**, 251 (2005), arXiv:quant-ph/0411036.
[10] S. Bravyi and J. Haah, Physical Review A **86**, 052329 (2012), arXiv:1209.2426.
[11] C. Jones, Physical Review A **87**, 042305 (2013), arXiv:1303.07847.
[12] C. J. Trout and K. R. Brown, International Journal of Quantum Chemistry **115**, 1296 (2015), arXiv:1501.01298.

[13] J. Haah, M. B. Hastings, D. Poulin, and D. Wecker, Quantum **1**, 31 (2017), arXiv:1703.07847.
[14] M. B. Hastings and J. Haah, Physical Review Letters **120**, 050504 (2018), arXiv:1709.03543.
[15] A. Krishna and J.-P. Tillich, arXiv e-prints (2018), arXiv:1811.08461.
[16] E. T. Campbell and M. Howard, Quantum **2**, 56 (2018), arXiv:1709.02214.
[17] X. Wang, M. M. Wilde, and Y. Su, arXiv e-prints (2018), arXiv:1812.10145.
[18] G. Vidal and R. Tarrach, Physical Review A **59**, 141 (1999), arXiv:quant-ph/9806094.
[19] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, Physical Review Letters **112**, 120401 (2014), arXiv:1209.3745.
[20] M. Horodecki and J. Oppenheim, International Journal of Modern Physics B **27**, 1345019 (2013), arXiv:1209.2162.
[21] F. G. S. L. Brandão and G. Gour, Physical Review Letters **115**, 070503 (2015), arXiv:1502.03149.
[22] B. Coecke, T. Fritz, and R. W. Spekkens, Information and Computation **250**, 59 (2016), arXiv:1409.5531.
[23] C. Napoli, T. R. Bromley, M. Cianciaruso, M. Piani, N. Johnston, and G. Adesso, Physical Review Letters **116**, 150502 (2016), arXiv:1601.03781.
[24] D. Stahlke, Physical Review A **90**, 022302 (2014), arXiv:1305.2186.
[25] R. Takagi and Q. Zhuang, Physical Review A **97**, 062337 (2018), arXiv:1804.04669.
[26] B. Regula, Journal of Physics A: Mathematical and Theoretical **51**, 045303 (2018), arXiv:1707.06298.
[27] D. Gross, Journal of Mathematical Physics **47**, 122107 (2006), arXiv:quant-ph/0602001.
[28] A. Mari and J. Eisert, Physical Review Letters **109**, 230503 (2012), arXiv:1208.3660.
[29] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, New Journal of Physics **14**, 113011 (2012), arXiv:1201.1256.
[30] V. Veitch, S. A. Hamed Mousavian, D. Gottesman, and J. Emerson, New Journal of Physics **16**, 013009 (2014), arXiv:1307.7171.
[31] M. Ahmadi, H. B. Dang, G. Gour, and B. C. Sanders, Physical Review A **97**, 062332 (2018), arXiv:1706.03828.
[32] N. Delfosse, C. Okay, J. Bermejo-Vega, D. E. Browne, and R. Raussendorf, New Journal of Physics **19**, 123024 (2017), arXiv:1610.07093.
[33] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Physical Review Letters **115**, 070501 (2015), arXiv:1503.07525.
[34] X. Wang, M. M. Wilde, and Y. Su, arXiv e-prints (2019), arXiv:1903.04483.
[35] N. Delfosse, P. Allard Guerin, J. Bian, and R. Raussendorf, Physical Review X **5**, 021003 (2015), arXiv:1409.5170.
[36] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega, Physical Review A **95**, 052334 (2017), arXiv:1511.08506.
[37] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurel, arXiv e-prints (2019), arXiv:1905.05374.
[38] M. Howard and E. Campbell, Physical Review Letters **118**, 090501 (2017), arXiv:1609.07488.
[39] H. J. García, I. L. Markov, and A. W. Cross, Quantum Info. Comput. **14**, 683 (2014), arXiv:1711.07848.
[40] S. Bravyi and D. Gosset, Physical Review Letters **116**, 250501 (2016), arXiv:1601.07601.
[41] S. Bravyi, G. Smith, and J. A. Smolin, Physical Review X **6**, 021043 (2016), arXiv:1506.01396.
[42] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, arXiv e-prints (2018), arXiv:1808.00128.
[43] P. Rall, D. Liang, J. Cook, and W. Kretschmer, arXiv e-prints (2019), arXiv:1901.09070.
[44] J. R. Seddon, "Channel magic - tools for calculating measures of magic for channels using MATLAB and CVX," `https://github.com/jamesrseddon/channel_magic` (2019).
[45] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. (Cambridge University Press, 2004).
[46] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," `http://cvxr.com/cvx` (2014).
[47] M. Heinrich and D. Gross, Quantum **3**, 132 (2019), arXiv:1807.10296.
[48] A. Jamiolkowski, Reports on Mathematical Physics **3**, 275 (1972).
[49] M.-D. Choi, Linear Algebra and its Applications **10**, 285 (1975).
[50] M. Jiang, S. Luo, and S. Fu, Physical Review A **87**, 022310 (2013).
[51] E. T. Campbell, Physical Review A **82**, 042314 (2010), arXiv:1007.1445.
[52] D. Fattal, T. S. Cubitt, Y. Yamamoto, S. Bravyi, and I. L. Chuang, arXiv e-prints (2004), arXiv:quant-ph/0406168.
[53] D. Gottesman and I. L. Chuang, Nature **402**, 390 (1999), arXiv:quant-ph/9908010.
[54] H. Pashayan, S. D. Bartlett, and D. Gross, arXiv e-prints , arXiv:1712.02806 (2017), arXiv:1712.02806.

[55] J. Dehaene and B. De Moor, Physical Review A **68**, 042318 (2003), arXiv:quant-ph/0304125.
[56] D. Gross and M. Van den Nest, Quantum Info. Comput. **8**, 263 (2008), arXiv:0707.4000.

Appendix A: Properties of robustness of the Choi state

Here we confirm that the robustness of the Choi state, $\mathcal{R}(\Phi)$ has the properties convexity and submultiplicativity under tensor product. We then give an example to show that it is not submultiplicative under composition.

Convexity: This follows immediately from convexity of robustness of magic. Consider a real linear combination of n-qubit channels: $\mathcal{E} = \sum_k q_k \mathcal{E}_k$. The Choi state for $\mathcal{E}$ is:

$$\Phi_{\mathcal{E}} = (\mathcal{E} \otimes \mathbb{1}_n) |\Omega_n\rangle\langle\Omega_n| \tag{A1}$$

$$= \sum_k q_k (\mathcal{E}_k \otimes \mathbb{1}_n) |\Omega_n\rangle\langle\Omega_n| = \sum_k q_k \Phi_{\mathcal{E}_k}, \tag{A2}$$

where in the last line we identified $(\mathcal{E}_k \otimes \mathbb{1}_n) |\Omega\rangle\langle\Omega|$ as the Choi state for $\mathcal{E}_k$. Then by convexity of robustness of magic:

$$\mathcal{R}(\Phi_{\mathcal{E}}) \le \sum_k |q_k| \mathcal{R}(\Phi_{\mathcal{E}_k}), \tag{A3}$$

which shows $\mathcal{R}(\Phi_{\mathcal{E}})$ is convex in $\mathcal{E}$.

Submultiplicativity under tensor product: The maximally entangled state $|\Omega_{n+m}\rangle^{AA'|BB'}$ as defined by equation (15) in the main text can be factored as $|\Omega_{n+m}\rangle^{AA'BB'} = |\Omega_n\rangle^{A|B} |\Omega_m\rangle^{A'|B'}$. So the Choi state for a channel $\mathcal{E}^{AA'} = \mathcal{E}^A \otimes \mathcal{E}'^{A'}$, where $\mathcal{E}^A$ and $\mathcal{E}'^{A'}$ are respectively n-qubit and m-qubit channels, can be written:

$$\Phi_{\mathcal{E}} = \left(\mathcal{E}^A \otimes \mathcal{E}'^{A'} \otimes \mathbb{1}_{n+m}\right) |\Omega_{n+m}\rangle\langle\Omega_{n+m}|^{AA'|BB'} \tag{A4}$$

$$= \left(\mathcal{E}^A \otimes \mathbb{1}_n\right) |\Omega_n\rangle\langle\Omega_n|^{A|B} \otimes \left(\mathcal{E}'^{A'} \otimes \mathbb{1}_m\right) |\Omega_m\rangle\langle\Omega_m|^{A'|B'} = \Phi_{\mathcal{E}^A} \otimes \Phi_{\mathcal{E}'^{A'}}. \tag{A5}$$

Then by submultiplicativity of robustness of magic for states, we have:

$$\mathcal{R}\left(\Phi_{\mathcal{E}^A \otimes \mathcal{E}'^{A'}}\right) \le \mathcal{R}(\Phi_{\mathcal{E}^A}) \mathcal{R}\left(\Phi_{\mathcal{E}'^{A'}}\right),$$

which is the desired property.

Failure of submultiplicativity under composition: Let $\mathcal{E}_1$ be the single-qubit Z-reset channel defined by Kraus operators $\{|0\rangle\langle 0|, |0\rangle\langle 1|\}$, and let $\mathcal{E}_2$ be the conditional channel defined by $\{|T\rangle\langle 0|, |1\rangle\langle 1|\}$, where $|T\rangle = T|+\rangle$. These channels respectively have Choi states $\Phi_{\mathcal{E}_1} = |0\rangle\langle 0| \otimes \frac{\mathbb{1}}{2}$, and $\Phi_{\mathcal{E}_2} = \frac{1}{2}(|T0\rangle\langle T0| + |11\rangle\langle 11|)$, with robustness of magic $\mathcal{R}(\Phi_{\mathcal{E}_1}) = 1$ and $\mathcal{R}(\Phi_{\mathcal{E}_2}) \approx 1.207$.

The composed channel $\mathcal{E}_2 \circ \mathcal{E}_1$ has a Kraus representation $\{|T\rangle\langle 0|, |T\rangle\langle 1|\}$, and so has a Choi state $\Phi_{\mathcal{E}_2 \circ \mathcal{E}_1} = |T\rangle\langle T| \otimes \frac{\mathbb{1}}{2}$, with $\mathcal{R}(\Phi_{\mathcal{E}_2 \circ \mathcal{E}_1}) \approx 1.414 > \mathcal{R}(\Phi_{\mathcal{E}_2}) \mathcal{R}(\Phi_{\mathcal{E}_1})$. So it is not the case that the robustness of the Choi state is submultiplicative under composition.

More intuitively, such counterexamples arise for channels $\mathcal{E}$ where the stabiliser state $|\phi_*\rangle$ that results in maximal final robustness $\mathcal{R}[(\mathcal{E} \otimes \mathbb{1}_n) |\phi_*\rangle\langle\phi_*|]$ is not the maximally entangled state $|\Omega_n\rangle$, as then we can always boost the output robustness by using a stabiliser-preserving operation to prepare $|\phi_*\rangle$ before applying $\mathcal{E}$.

Appendix B: Properties of channel robustness

Faithfulness: Suppose $\mathcal{E}$ is an n-qubit CPTP map. There are two cases:

(i) $\mathcal{E} \in \mathrm{SP}_{n,n}$. In this case, $\Phi_\mathcal{E}$ is itself a mixed stabiliser state, and since $\mathcal{E}$ is trace-preserving it satisfies $\mathrm{Tr}(\Phi_\mathcal{E}) = \mathbb{1}_n/2^n$. So $\Phi_\mathcal{E}$ is already trivially a decomposition of the correct form, with $p = 0$, so that $R_*(\mathcal{E}) = 1 + 2p = 1$.

(ii) $\mathcal{E} \notin \mathrm{SP}_{n,n}$. Then by faithfulness of robustness of magic (Lemma 4.2 in the main text), $\Phi_\mathcal{E}$ has $\mathcal{R}(\Phi_\mathcal{E}) > 1$. Since the definition of $\mathcal{R}_*$ is a restriction of $\mathcal{R}(\Phi)$, it must be the case that $\mathcal{R}(\Phi_\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})$. Therefore $\mathcal{R}_*(\mathcal{E}) > 1$.

Convexity:

Suppose we have a set of Choi states $\Phi_{\mathcal{E}_j}$ corresponding to channels $\mathcal{E}_j$, with optimal decompositions:

$$\Phi_{\mathcal{E}_j} = (1 + p_j)\rho_{j+} - p_j\rho_{j-}, \tag{A1}$$

where each $\rho_{j\pm}$ separately satisfies the condition $\mathrm{Tr}_A(\rho_\pm) = \frac{\mathbb{1}_n}{2^n}$, so that $\mathcal{R}_*(\mathcal{E}_j) = 1 + 2p_j$. Now take a real linear combination of such channels:

$$\mathcal{E} = \sum_i q_i\mathcal{E}_i = \sum_{j \in P} q_j\mathcal{E}_j + \sum_{k \in N} q_k\mathcal{E}_k, \tag{A2}$$

where P is the set of indices such that $q_j \geq 0$, and N is the set such that $q_k < 0$. We assume that $\sum_i q_i = 1$ so that the trace of $\mathrm{Tr}(\Phi_\mathcal{E}) = 1$. Then the corresponding Choi state for channel $\mathcal{E}$ is:

$$\Phi_\mathcal{E} = \sum_{j \in P} q_j\Phi_{\mathcal{E}_j} - \sum_{k \in N} |q_k|\Phi_{\mathcal{E}_k} \tag{A3}$$

$$= \sum_{j \in P} q_j\big[(1 + p_j)\rho_{j+} - p_j\rho_{j-}\big] - \sum_{k \in N} |q_k|\big[(1 + p_k)\rho_{k+} - p_k\rho_{k-}\big] \tag{A4}$$

$$= \left(\sum_{j \in P} q_j(1 + p_j)\rho_{j+} + \sum_{k \in N} |q_k|p_k\rho_{k-}\right) - \left(\sum_{j \in P} q_j p_j\rho_{j-} + \sum_{k \in N} |q_k|(1 + p_k)\rho_{k+}\right). \tag{A5}$$

Note that the terms inside the brackets all have positive coefficients, hence we can interpret as non-normalised mixtures over stabiliser states. To normalise them we can define:

$$\widetilde{\rho}_+ = \frac{\sum_{j \in P} q_j(1 + p_j)\rho_{j+} + \sum_{k \in N} |q_k|p_k\rho_{k-}}{\sum_{j \in P} q_j(1 + p_j) + \sum_{k \in N} |q_k|p_k}, \tag{A6}$$

and

$$\widetilde{\rho}_- = \frac{\sum_{j \in P} q_j p_j\rho_{j-} + \sum_{k \in N} |q_k|(1 + p_k)\rho_{k+}}{\sum_{j \in P} q_j p_j + \sum_{k \in N} |q_k|(1 + p_k)}. \tag{A7}$$

Then writing:

$$\widetilde{p} = \sum_{j \in P} q_j p_j + \sum_{k \in N} |q_k|(1 + p_k), \tag{A8}$$

one can check that:

$$1 + \widetilde{p} = \sum_{j \in P} q_j(1 + p_j) + \sum_{k \in N} |q_k|p_k. \tag{A9}$$

This allows us to rewrite the Choi state as: $\Phi_\mathcal{E} = (1 + \widetilde{p})\widetilde{\rho}_+ - \widetilde{p}\widetilde{\rho}_-$. Since $\widetilde{\rho}_\pm$ are convex mixtures over stabiliser states satisfying $\mathrm{Tr}_A(\rho_{j\pm}) = \frac{\mathbb{1}_n}{2^n}$, they must satisfy the same condition. We also know that $\widetilde{p} \geq 0$, so it is clear that the decomposition is in the form required for the definition of

$\mathcal{R}_*$, except that it is not necessarily optimised to minimise $1 + 2\widetilde{p}$. So we have:

$$\mathcal{R}_*\left(\sum_j q_j \mathcal{E}_j\right) \leq 1 + 2\widetilde{p} = \sum_{j \in P} q_j(1 + p_j) + \sum_{k \in N} |q_k| + \sum_{j \in P} q_j p_j + \sum_{k \in N} |q_k|(1 + p_k) \tag{A10}$$

$$= \sum_{j \in P} |q_j|(1 + 2p_j) + \sum_{k \in N} |q_k|(1 + 2p_k) \tag{A11}$$

$$= \sum_i |q_i| \mathcal{R}_*(\mathcal{E}_j), \tag{A12}$$

which gives us the required result.

Invariance under tensor with identity: In Section 5 of the main text we saw that $\mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}) \leq \mathcal{R}_*(\mathcal{E}^A)$. We now complete the proof that $\mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}) = \mathcal{R}_*(\mathcal{E}^A)$ by showing that $\mathcal{R}_*(\mathcal{E}^A) \leq \mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1})$.

Consider an optimal decomposition for $\Phi_{\mathcal{E}^A \otimes \mathbb{1}_m} = (1 + p')\rho'_+ - p'\rho'_-$, such that $\mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}_m) = 1 + 2p'$, where $\mathrm{Tr}_{AA'}(\rho_\pm) = \mathbb{1}_{n+m}/2^{n+m}$. Here we do not assume that $\rho'_\pm$ are products across the partition $AB|A'B'$, as was the case in equation (34) in the main text. However, we have just seen that $\Phi_{\mathcal{E}^A \otimes \mathbb{1}_m}$ can be written as a product, so that by tracing out systems $A'B'$ we obtain:

$$\Phi_{\mathcal{E}^A} = (1 + p') \mathrm{Tr}_{A'B'}(\rho'_+) - p' \mathrm{Tr}_{A'B'}(\rho'_-). \tag{A13}$$

Partial trace of a stabiliser state remains a stabiliser state, so this is a stabiliser decomposition. We just need to check that the partial trace condition holds, so we want to show:

$$\mathrm{Tr}_A(\mathrm{Tr}_{A'B'}(\rho'_\pm)) = \mathrm{Tr}_{AA'B'}(\rho'_\pm) = \frac{\mathbb{1}_n}{2^n}, \tag{A14}$$

but this is clearly the case from the fact that $\rho'_\pm$ were constrained such that $\mathrm{Tr}_{AA'}(\rho_\pm) = \mathbb{1}_{n+m}/2^{n+m}$. Hence again we have a valid, not necessarily optimal decomposition and:

$$\mathcal{R}_*(\mathcal{E}^A) \leq 1 + 2p' = \mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}^B). \tag{A15}$$

Combining with the inequality $\mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}^B) \leq \mathcal{R}_*(\mathcal{E}^A)$, shown in the main text, we obtain the equality:

$$\mathcal{R}_*(\mathcal{E} \otimes \mathbb{1}) = \mathcal{R}_*(\mathcal{E}) = \mathcal{R}_*(\mathbb{1} \otimes \mathcal{E}). \tag{A16}$$

Appendix C: Optimisation problem for channel robustness

In Howard and Campbell [38], the optimisation problem for calculating robustness of magic for states was cast as follows:

$$\textbf{minimise} \quad \|\vec{q}\|_1$$
$$\textbf{subject to} \quad A\vec{q} = \vec{b},$$

where $\vec{q}$ is a vector of coefficients, $\vec{b}$ is the vector of Pauli expectation values for the target state $\Phi_{\mathcal{E}}$, and A is a matrix whose columns are the Pauli vectors for the stabiliser states. For n-qubit channels, we have $2n$-qubit Choi states, so the number of generalised Paulis is $N_P = 4^{2n}$, and the number of stabiliser states is $N_S = 2^{2n} \prod_{j=1}^{2n}(2^j + 1)$ [38]. Then $\vec{b}$ has N_P entries, $\vec{q}$ has N_S entries, and the dimension of A is $(N_P \times N_S)$. From this construction we can recover optimal decompositions of the form: $\Phi_{\mathcal{E}} = \sum_j q_j |\phi_j\rangle\langle\phi_j|$, where $\sum_j q_j = 1$ and $|\phi_j\rangle$ are the pure stabiliser

states.

We want to restrict the problem to decompositions of the form:

$$\Phi_{\mathcal{E}} = (1+p)\rho_+ - p\rho_-, \tag{A1}$$

where $p \geq 0$ and $\rho_\pm$ correspond to trace-preserving channels, and can in general be mixed. Rather than enumerating all the extreme points of the set of stabiliser states corresponding to maps in $\mathrm{SP}_{n,n}$, it is more convenient to retain the same A matrix and modify the constraints. We still need to start from a finite set of extreme points, i.e. pure stabiliser states, so first rewrite as:

$$\Phi_{\mathcal{E}} = \sum_j q_{j_+}\rho_j + \sum_j q_{j_-}\rho_j = \sum_j p_{j_+}\rho_j - \sum_j p_{j_-}\rho_j, \tag{A2}$$

where q_{j_+} are the positive quasiprobabilities, q_{j_-} are the negative quasiprobabilities, and $p_{j_\pm} = |q_{j_\pm}|$. In the Pauli vector picture we can write this as $\vec{b} = A\vec{p}_+ - A\vec{p}_-$, where all the entries of $\vec{p}_\pm$ are non-negative. We define a new variable vector $\vec{p}$ which will have twice the length of the previous $\vec{q}$, i.e. $2N_S$ entries:

$$\vec{p} = \begin{pmatrix} \vec{p}_+ \\ \vec{p}_{-,} \end{pmatrix} \tag{A3}$$

and define a new $(N_P \times 2N_S)$ matrix A' in block form, $A' = \begin{pmatrix} A & -A \end{pmatrix}$. Then we have:

$$A'\vec{p} = \begin{pmatrix} A & -A \end{pmatrix} \begin{pmatrix} \vec{p}_+ \\ \vec{p}_- \end{pmatrix} = A\vec{p}_+ - A\vec{p}_- = \vec{b}. \tag{A4}$$

So now we need to minimise $\|\vec{p}\|_1 = \sum_j p_j$ subject to $A'\vec{p} = \vec{b}$ and $\vec{p} \geq 0$.

Next, we need the trace-preserving condition. Provided $\mathcal{E}$ is CPTP, if one part of the decomposition is trace-preserving, then the other will be as well, so we only need enforce the constraint on one of ρ_+ or ρ_-. Assume that we check ρ_+. The condition for a Choi state $\Phi^{AB} = \mathcal{E}^A \otimes \mathbb{1}^B(|\Omega\rangle\langle\Omega|^{AB})$ to be trace-preserving is:

$$\mathrm{Tr}_A(\Phi^{AB}) = \frac{\mathbb{1}}{d}, \tag{A5}$$

where d is the dimension of the subsystem. We need to convert this to a constraint on the vector $\vec{b}_+$ corresponding to ϕ_+, which is given by $\vec{b}_+ = A\vec{p}_+$. First, note that all Paulis are traceless except for the identity $P_0 = \mathbb{1}$, so for the maximally mixed state:

$$\langle P_j \rangle = \mathrm{Tr}\left(P_j \frac{\mathbb{1}}{d}\right) = \frac{\mathrm{Tr}(P_j)}{d} = \delta_{j,0}, \tag{A6}$$

so if the first entry in a Pauli vector is always $\langle \mathbb{1} \rangle$, the maximally mixed state has Pauli vector:

$$\vec{b}_B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{A7}$$

where $\vec{0}$ is the zero vector. However, we need this to hold just for the reduced state on B rather than the full Pauli vector. Consider that if the whole state is written $\Phi^{AB} = \sum_{j,k} r_{j,k} P_j \otimes P_k$ for some set of coefficients $r_{j,k}$, then the expectation values are given by:

$$\langle P_l \otimes P_m \rangle = \sum_{j,k} r_{j,k} \mathrm{Tr}(P_l P_j \otimes P_m P_k) = \sum_{j,k} r_{j,k} d^2 \delta_{j,l}\delta_{m,k} = d^2 r_{l,m}. \tag{A8}$$

The reduced state is:

$$\text{Tr}_A(\Phi^{AB}) = \sum_{j,k} r_{j,k} \, \text{Tr}_A[P_j \otimes P_k] = \sum_{j,k} r_{j,k} d\delta_{j,0} P_k = d \sum_k r_{0,k} P_k. \tag{A9}$$

and the entries of the reduced Pauli vector will be:

$$\langle P_m \rangle = d \sum_k r_{0,k} \, \text{Tr}\{P_m P_k\} = d^2 r_{0,m} = \langle P_0 \otimes P_m \rangle. \tag{A10}$$

So for condition (A5) to hold for the reduced state on B, we combine equations (A6) and (A10) to get:

$$\langle P_m \rangle = \langle P_0 \otimes P_m \rangle = \delta_{m,0}. \tag{A11}$$

That is, we just need to look at the entries of $\vec{b}_+$ corresponding to Paulis of the form $\mathbb{1} \otimes P_j$. These should all be zero except the first entry, which corresponds to $\langle \mathbb{1} \otimes \mathbb{1} \rangle$. Note that $\vec{b}_+ = A\vec{p}_+$ will in general not be normalised, but this does not matter, since we are only interested in whether or not entries are zero. We can use a binary matrix M to pick out the values of interest. As an example we consider the two-qubit case, and assume that the entries are ordered as:

$$\vec{b}_+ = \begin{pmatrix} \langle \mathbb{1} \otimes \mathbb{1} \rangle \\ \langle \mathbb{1} \otimes X \rangle \\ \langle \mathbb{1} \otimes Y \rangle \\ \langle \mathbb{1} \otimes Z \rangle \\ \langle X \otimes \mathbb{1} \rangle \\ \vdots \\ \langle Z \otimes Z \rangle \end{pmatrix}. \tag{A12}$$

Here, we are only interested in the 2nd, 3rd and 4th entries. We form a new vector $\vec{c}$ by left multiplying with M:

$$\vec{c} = M\vec{b}_+ = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \vec{b}_+ = \begin{pmatrix} \langle \mathbb{1} \otimes X \rangle \\ \langle \mathbb{1} \otimes Y \rangle \\ \langle \mathbb{1} \otimes Z \rangle \end{pmatrix}. \tag{A13}$$

Then the condition we need is just $\vec{c} = 0$. To convert this to a condition on the $2N_S$-entry variable $\vec{p} = \begin{pmatrix} \vec{p}_+ \\ \vec{p}_- \end{pmatrix}$, we first pad A with zeroes: $A_+ = \begin{pmatrix} A & \bar{0} \end{pmatrix}$, where $\bar{0}$ is the $(N_P \times N_S)$ zero matrix. We then have:

$$\vec{b}_+ = A\vec{p}_+ = A\vec{p}_+ + \bar{0}\vec{p}_- = \begin{pmatrix} A & \bar{0} \end{pmatrix} \begin{pmatrix} \vec{p}_+ \\ \vec{p}_- \end{pmatrix} = A_+\vec{p}, \tag{A14}$$

so that $\vec{c} = M\vec{b}_+ = MA_+\vec{p}$. Therefore, our condition for trace-preserving ρ_+ is $MA_+\vec{p} = 0$. We can therefore specify the new optimisation problem as:

$$\begin{aligned} \textbf{minimise} \quad & \|\vec{p}\|_1 = \sum_j p_j \\ \textbf{subject to} \quad & A'\vec{p} = \vec{b}, \\ & \vec{p} \geq 0, \\ & MA_+\vec{p} = 0 \end{aligned}$$

where $A' = \begin{pmatrix} A & -A \end{pmatrix}$, and $A_+ = \begin{pmatrix} A & \vec{0} \end{pmatrix}$, with A and $\vec{b}$ having the same definitions as previously, $\overline{0}$ is the zero matrix with dimension the same as A, and with M being the binary matrix that picks out the $\langle \mathbb{1} \otimes P_j \rangle$ entries from the vector $A_+\vec{p}$. Most of this is straightforward to implement. The step that requires some care is in correctly constructing the matrix M, as it will depend on the choice of ordering of Pauli operators in the construction of A and $\vec{b}$. If the B subsystem has n qubits, then we will need to constrain $4^n - 1$ non-trivial $\langle \mathbb{1} \otimes P_j \rangle$ expectation values to zero, so M should have dimension $((4^n - 1) \times N_P)$. If the Paulis are ordered as in the example given above for 2-qubit Choi states, then the construction is just $M = \begin{pmatrix} \vec{0} & \mathbb{1}' & \vec{0} & \cdots & \vec{0} \end{pmatrix}$, where $\mathbb{1}'$ is the $((4^n - 1) \times (4^n - 1))$ identity, and $\vec{0}$ denotes a column of zeroes. We have implemented this linear program in MATLAB, using the convex optimisation package CVX [46], and have made the code available from the repository Ref. [44].

Appendix D: Properties of magic capacity

Faithfulness: For any n-qubit stabiliser-preserving CPTP channel Λ, if $\rho \in \text{STAB}_{2n}$ is a stabiliser state, then $(\Lambda \otimes \mathbb{1}_n)\rho$ is also a stabiliser state. So by the faithfulness of robustness of magic, $\mathcal{R}((\Lambda \otimes \mathbb{1}_n)\rho) = 1$ for any input stabiliser state $\rho \in \text{STAB}_{2n}$, and $\mathcal{C}(\Lambda) = 1$.

Suppose instead that $\mathcal{E}$ is non-stabiliser-preserving, but still CPTP. Then there exists at least one stabiliser state $\rho \in \text{STAB}_{2n}$ such that $(\mathcal{E} \otimes \mathbb{1})\rho$ is a normalised state, but not a stabiliser state. Then by faithfulness of $\mathcal{R}$ when applied to states, $\mathcal{R}((\mathcal{E} \otimes \mathbb{1})\rho) > 1$, and so $\mathcal{C}(\mathcal{E}) > 1$.

Convexity: Suppose we have a real linear combination of n-qubit CPTP maps $\mathcal{E}_k$:

$$\mathcal{E} = \sum_k q_k \mathcal{E}_k. \tag{A1}$$

There exists some optimal stabiliser state ρ_* that achieves $\mathcal{C}(\mathcal{E}) = \mathcal{R}(\mathcal{E} \otimes \mathbb{1}(\rho_*))$. Then

$$\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)\rho_*) = \mathcal{R}\left(\sum_k q_k[(\mathcal{E}_k \otimes \mathbb{1}_n)\rho_*]\right) \tag{A2}$$

$$\leq \sum_k |q_k| \mathcal{R}((\mathcal{E}_k \otimes \mathbb{1}_n)\rho_*), \tag{A3}$$

where the last line follows by convexity of the robustness of magic. But each robustness $\mathcal{R}((\mathcal{E}_k \otimes \mathbb{1}_n)\rho_*)$ can be no larger than $\mathcal{C}(\mathcal{E}_k)$. So we have:

$$\mathcal{C}\left(\sum_k q_k \mathcal{E}_k\right) \leq \sum_k |q_k| \mathcal{C}(\mathcal{E}_k). \tag{A4}$$

Appendix E: Calculating monotones for diagonal channels

1. Reducing the problem size

As mentioned earlier, the size of the optimisation problem for calculating our monotones (as well as $\mathcal{R}(\Phi_\mathcal{E})$) quickly becomes prohibitively large for n-qubit states, since the number of stabiliser states increases super-exponentially with n (Table II). The issue is even worse than it first appears, since for an n-qubit channel we must in general consider $2n$-qubit stabiliser states. Direct calculation of either monotone is impractical for n-qubit channels with $n > 2$. This difficulty is aggravated when calculating the capacity as in principle we have to repeat the optimisation for every $(\mathcal{E} \otimes \mathbb{1}_n)|\phi\rangle\langle\phi|$ such that $|\phi\rangle \in \text{STAB}_{2n}$. In some cases we can ameliorate these problems by

n	N_S
1	6
2	60
3	1,080
4	36,720
5	2,423,520
6	315,057,600

TABLE II. Number of pure stabiliser states N_S for number of qubits n.

looking for Clifford gates that commute with the channel of interest. Here we consider the case where $\mathcal{E}$ is a diagonal channel, meaning it has a Kraus representation where each Kraus operator is diagonal in the computational basis. This of course includes diagonal unitaries as a special case. One could likely reduce the problem size further by exploiting symmetries of channels using techniques similar to those used in Ref. [47], but we will not consider this strategy here.

It is straightforward to see how the problem can be simplified for calculating $\mathcal{R}(\Phi_\mathcal{E})$ and $\mathcal{R}_*(\mathcal{E})$. If $\mathcal{E}$ is diagonal, the operation $\mathcal{E} \otimes \mathbb{1}_n$ commutes with any sequence of CNOTs targeted on the last n qubits. But the maximally entangled state $|\Omega_n\rangle$ can be written:

$$|\Omega_n\rangle = U_C(|+\rangle^{\otimes n} \otimes |0\rangle^{\otimes n}).\qquad(A1)$$

Here $U_C = \otimes_{j=1}^n U_j$, where U_j is the CNOT controlled on qubit j and targeted on qubit $n+j$. By the monotonicity of robustness of magic, we immediately see that:

$$\mathcal{R}(\Phi_\mathcal{E}) = \mathcal{R}[(\mathcal{E} \otimes \mathbb{1}_n)\,|\Omega_n\rangle\langle\Omega_n|] = \mathcal{R}\big[\mathcal{E}\big(|+\rangle\langle+|^{\otimes n}\big) \otimes |0\rangle\langle0|^{\otimes n}\big] = \mathcal{R}\big[\mathcal{E}\big(|+\rangle\langle+|^{\otimes n}\big)\big].\qquad(A2)$$

For the channel robustness we would like to decompose $\mathcal{E}\big(|+\rangle\langle+|^{\otimes n}\big)$ in terms of states $\rho_\pm \in$ STAB$_n$, but need to take care that the trace condition $\text{Tr}_A(\rho'_\pm) = \mathbb{1}_n/2^n$ is satisfied for the equivalent $2n$-qubit Choi states $\rho'_\pm$. In Appendix E2 we show that the criterion is satisfied provided all diagonal elements of $\rho_\pm$ are equal to $1/2^n$. So for diagonal channels we can write:

$$\mathcal{R}_*(\mathcal{E}) = \min_{\rho_\pm \in \text{STAB}_n}\left\{1 + 2p : (1+p)\rho_+ - p\rho_- = \mathcal{E}(|+\rangle\langle+|^{\otimes n}),\ p \geq 0, \langle x|\,\rho_\pm\,|x\rangle = \frac{1}{2^n}, \forall x\right\}.\quad(A3)$$

So calculation of $R_*(\mathcal{E})$ and $\mathcal{R}(\Phi_\mathcal{E})$ is tractable up to five qubits provided $\mathcal{E}$ is diagonal. We will see below in Section E3 that this is also true for the magic capacity.

2. Trace condition for diagonal channels

Consider that the Choi state for a diagonal channel has a decomposition

$$\Phi_\mathcal{E} = U_C\big(\mathcal{E}(|+\rangle\langle+|^{\otimes n}) \otimes |0\rangle\langle0|^{\otimes n}\big)U_C^\dagger = (1+p)\rho_+ - p\rho_-,\qquad(A4)$$

where $U_C = \otimes_{j=1}^n U_j$ is the tensor product of CNOTs U_j that are controlled on the jth qubit and targeted on the $n+j$th. Then

$$\mathcal{E}(|+\rangle\langle+|^{\otimes n}) \otimes |0\rangle\langle0|^{\otimes n} = (1+p)\rho'_+ - p\rho'_-,\qquad(A5)$$

where $\rho'_\pm$ are still stabiliser states since U_C is Clifford. Now consider the stabiliser-preserving channel $\mathbb{1}_n \otimes \Lambda$ that resets the last n qubits to $|0\rangle\langle0|^{\otimes n}$. Applying this to both sides of equation

(A5) we get a new decomposition

$$\mathcal{E}(|+\rangle\langle+|^{\otimes n}) \otimes |0\rangle\langle0|^{\otimes n} = (1+p)\rho''_+ \otimes |0\rangle\langle0|^{\otimes n} - p\rho''_- \otimes |0\rangle\langle0|^{\otimes n}. \tag{A6}$$

Then referring back to equation (A4), we obtain $\rho_\pm = U_C(\rho''_+ \otimes |0\rangle\langle0|^{\otimes n})U_C^\dagger$. So, the trace-preserving condition becomes:

$$\frac{\mathbb{1}_n}{2^n} = \text{Tr}_A(\rho_\pm) = \text{Tr}_A\left(U_C(\rho''_+ \otimes |0\rangle\langle0|^{\otimes n})U_C^\dagger\right) \tag{A7}$$

$$= \sum_x \langle x|^A U_C(\rho''_+ \otimes |0\rangle\langle0|^{\otimes n})U_C^\dagger |x\rangle^A, \tag{A8}$$

where $|x\rangle$ are the computational basis states on subsystem A. Recalling that U_C can be written as a tensor product of CNOTs $U_C = \otimes_{j=1}^n U_j$ one can check that this equation can be written:

$$\frac{\mathbb{1}_n}{2^n} = \sum_x \langle x| \rho''_\pm |x\rangle |x\rangle\langle x|. \tag{A9}$$

Therefore, the decomposition corresponds to a pair of trace-preserving channels provided that all diagonal elements of $\rho''_\pm$ are equal to $1/2^n$.

For a given diagonal channel, there always exists a decomposition that satisfies these conditions and has ℓ_1-norm equal to the channel robustness as defined for the full Choi state. We do not give the full proof here, but sketch the argument. Given any diagonal channel $\mathcal{E}$ decomposition of the full Choi state $\Phi_\mathcal{E} = (1+p)\rho_+ - p\rho_-$ satisfying the trace condition, one can always find a new decomposition $\Phi_\mathcal{E} = (1+p)\Lambda(\rho_+) - p\Lambda(\rho_-)$ where $\Lambda(\rho_\pm)$ still satisfy $\text{Tr}_A(\Lambda(\rho_\pm))$, but are now the Choi states for diagonal channels. The map Λ used to obtain this decomposition is in effect an error correction circuit that takes general stabiliser Choi states to the subspace corresponding to the diagonal channels. Specifically, we note that the Choi states for diagonal maps $\mathcal{T}$ have the general form:

$$\Phi_\mathcal{T} = \frac{1}{2^n} \sum_{j,k} c_{j,k} |j\rangle^A |j\rangle^B \langle k|^A \langle k|^B. \tag{A10}$$

In general $c_{j,k}$ can be complex or zero, but terms on the diagonal are constrained. In particular, trace-preserving diagonal channels cannot change the weight of particular computational basis states, so the probability distribution for computational basis states will be the same as for $|\Omega\rangle$:

$$\langle p,q| \Phi_\mathcal{T} |p,q\rangle = \frac{1}{2^n}\delta_{p,q}. \tag{A11}$$

The circuit Λ is defined by the following steps. For each j from 1 to n:

1. Perform a parity measurement $(Z \otimes Z)$ between qubits j and $n+j$.

2. If even parity (+1 outcome), do nothing. If odd parity (-1 outcome), perform an X gate on qubit j.

This stabiliser-preserving channel leaves Choi states for diagonal maps (and crucially, the target Choi state $\Phi_\mathcal{E}$) invariant, but updates general Choi states to have the form (A10). One can check that the circuit preserves the property $\text{Tr}_A(\rho_\pm) = \mathbb{1}_n/2^n$. We then obtain a decomposition in the desired form:

$$\Lambda(\Phi_\mathcal{E}) = \Phi_\mathcal{E} = (1+p)\Lambda(\rho_+) - p\Lambda(\rho_-) \tag{A12}$$

Where $\Lambda(\rho_\pm) = (\mathcal{E}_\pm \otimes \mathbb{1}) |\Omega\rangle\langle\Omega|$ are Choi states for n-qubit diagonal channels $\mathcal{E}_\pm$. But as described above, the CNOT sequence U_C commutes with diagonal channels acting on the first n qubits, so we can obtain n-qubit representatives of these channels:

$$\mathcal{E}_\pm(|+\rangle\langle+|^{\otimes n}) \otimes |0\rangle\langle 0|^{\otimes n} = U_C((\mathcal{E}_\pm \otimes \mathbb{1}) |\Omega\rangle\langle\Omega|)U_C^\dagger. \tag{A13}$$

Discarding the last n qubits we obtain the desired n-qubit decomposition:

$$\mathcal{E}(|+\rangle\langle+|^{\otimes n}) = (1 + p)\mathcal{E}_+(|+\rangle\langle+|^{\otimes n}) - p\mathcal{E}_-(|+\rangle\langle+|^{\otimes n}). \tag{A14}$$

3. Magic capacity in the affine space picture

In this section we will make use of the formalism due to Dehaene and De Moor, in which stabiliser states are cast in terms of affine spaces and quadratic forms over binary vectors [55, 56], to prove the following theorem:

Theorem A1 (Capacity for diagonal operations). *Suppose the n-qubit channel $\mathcal{E}_D$ is diagonal. Let:*

$$|\mathcal{K}\rangle = \frac{1}{|\mathcal{K}|^{1/2}} \sum_{x \in \mathcal{K}} |x\rangle, \tag{A15}$$

where $x \in \mathbb{F}_2^n$ are binary vectors and $\mathcal{K} \subseteq \mathbb{F}_2^n$ is an affine space. Then:

$$\mathcal{C}(\mathcal{E}_D) = \max_{\mathcal{K}} \mathcal{R}(\mathcal{E}_D(|\mathcal{K}\rangle\langle\mathcal{K}|)). \tag{A16}$$

That is, given an n-qubit channel $\mathcal{E}$, provided the channel is diagonal, the capacity $\mathcal{C}(\mathcal{E})$ may be calculated by optimisation over only the n-qubit states $|\mathcal{K}\rangle$ as defined in equation (A15), rather than over all $2n$-qubit stabiliser states.

We first review the formalism of Ref. [55]. Computational basis states $|x\rangle$ can be labelled by binary column vectors $x = (x_1, \ldots, x_n)^T \in \mathbb{F}_2^n$, so that $x_j \in \{0, 1\}$ relates to the jth qubit. Any pure n-qubit stabiliser state may be written:

$$|\mathcal{K}, q, d\rangle = \frac{1}{|\mathcal{K}|^{1/2}} \sum_{x \in \mathcal{K}} i^{d^T x}(-1)^{q(x)} |x\rangle, \tag{A17}$$

where $\mathcal{K} \subseteq \mathbb{F}_2^n$ is an affine space, d is some fixed binary vector, and $q(x)$ has the form:

$$q(x) = x^T Q x + \lambda^T x. \tag{A18}$$

Here Q is a binary, strictly upper triangular matrix, λ is a vector, and addition is modulo 2. Conversely, any state that can be written in this way is a stabiliser state.

An affine space $\mathcal{K}$ is a linear subspace $\mathcal{L}$ shifted by some constant binary vector h, modulo 2: $\mathcal{K} = \mathcal{L} + h$. Every affine space is related in this way to exactly one linear subspace, and the dimension $k = \dim(\mathcal{K})$ of an affine space means the dimension of the corresponding subspace. Instead of enumerating all elements of an affine space, we can specify it by a shift vector h and an $n \times k$ matrix where each column is one of the generators of the corresponding linear space:

$$G = \begin{pmatrix} \vec{g}_1 & \vec{g}_2 & \cdots & \vec{g}_k \end{pmatrix} = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,k} \\ \vdots & \vdots & & \vdots \\ g_{j,1} & g_{j,2} & \cdots & g_{j,k} \\ \vdots & \vdots & & \vdots \\ g_{n,1} & g_{n,2} & \cdots & g_{n,k} \end{pmatrix}. \tag{A19}$$

We have freedom in our choice of k independent generators, and we can transform between equivalent generating sets by adding any two columns of G. We are also free to swap any two columns. A general transform between generating sets can therefore be represented by an invertible matrix S of dimension $k \times k$, multiplying on the right $G \longrightarrow GS$.

Any non-trivial linear transformation of the affine space can be fully specified by the transformation of the generators and the shift vector. In particular, we can represent the action of a single CNOT by multiplication on the left by a matrix C. If the CNOT has control qubit j and target qubit k, then C has 1s on the diagonal, a 1 in the jth element of the kth row, and zeroes everywhere else. A sequence for a $2n$-qubit system, in which CNOTs are always controlled on the first n qubits, and targeted on the last n qubits can be represented in block form:

$$C = \begin{pmatrix} \mathbb{1} & 0 \\ M & \mathbb{1} \end{pmatrix}, \tag{A20}$$

where each block has dimension $n \times n$, and M can be any binary matrix. We use this formalism to prove the following lemma, which leads directly to Theorem A1:

Lemma A1 (Equivalences for diagonal channels). *Suppose $\mathcal{E}_D$ is a diagonal CPTP channel. Then:*

1. *All input stabiliser states with the same affine space $\mathcal{K}$ result in the same final robustness:*

$$\mathcal{R}((\mathcal{E}_D \otimes \mathbb{1}) |\mathcal{K}, q, d\rangle\langle\mathcal{K}, q, d|) = \mathcal{R}\big((\mathcal{E}_D \otimes \mathbb{1}) |\mathcal{K}, q', d'\rangle\langle\mathcal{K}, q', d'|\big), \quad \forall q, q', d, d'. \tag{A21}$$

2. *Given a $2n$-qubit state $|\phi\rangle \in \mathrm{STAB}_{2n}$, there exists some n-qubit $|\phi'\rangle \in \mathrm{STAB}_n$ such that:*

$$\mathcal{R}((\mathcal{E}_D \otimes \mathbb{1}_n) |\phi\rangle\langle\phi|) = \mathcal{R}\big(\mathcal{E}_D\big(|\phi'\rangle\langle\phi'|\big)\big). \tag{A22}$$

Proof. We first prove statement 1. Since robustness of magic is invariant under Clifford unitaries, we need to show that there exists a Clifford unitary U that converts $(\mathcal{E}_D \otimes \mathbb{1}) |\mathcal{K}, q, d\rangle\langle\mathcal{K}, q, d|$ to $(\mathcal{E}_D \otimes \mathbb{1}) |\mathcal{K}, q', d'\rangle\langle\mathcal{K}, q', d'|$. A suitable choice for U is one such that $U |\phi_{\mathcal{K},q,d}\rangle = |\phi_{\mathcal{K},q',d'}\rangle$, and, crucially, that commutes with the channel $\mathcal{E}_D$. Since $\mathcal{E}_D$ is given to be diagonal, any diagonal Clifford U will suffice. The affine space $\mathcal{K}$ remains unchanged, so we only need show there is always a diagonal Clifford that maps $q \to q'$ and $d \to d'$ for any q, q', d and d'. That this is always possible is perhaps already evident from Ref. [55], but for completeness we give the argument here.

We can convert d to d' using appropriately chosen S_j gates, meaning the gate $\mathrm{diag}(1, i)$ acting on the jth qubit. Consider the action of S_j on a basis vector:

$$S_j |x\rangle = \begin{cases} |x\rangle \text{ if } x_j = 0 \\ i |x\rangle \text{ if } x_j = 1 \end{cases}. \tag{A23}$$

If we define basis vector e_j so that it has 1 in the jth position and zeroes elsewhere, we can write the action of S_j as:

$$S_j |x\rangle = i^{e_j^T x} |x\rangle. \tag{A24}$$

Note that the form of this equation is independent of the value of x, so we can write:

$$S_j |\phi_{\mathcal{K},q,d}\rangle = \frac{1}{|\mathcal{K}|^{1/2}} \sum_{x\in\mathcal{K}} i^{d^T x}(-1)^{q(x)} S_j |x\rangle \tag{A25}$$

$$= \sum_{x\in\mathcal{K}} i^{(d^T + e_j^T)x}(-1)^{q(x)} S_j |x\rangle . \tag{A26}$$

So, we can flip any bit of d by applying the correct S gate. The quadratic form $q(x)$ is left unchanged.

Now consider $q(x) = x^T Q x + \lambda^T x$, which we must convert to some other $q'(x) = x^T Q' x + \lambda'^T x$. We can use the same trick as above to convert any λ to any other λ', by replacing S_j with the Z_j gate, i.e. $\mathrm{diag}(1,-1)$ acting on the jth qubit. For Q we can use the controlled-Z gate between the jth and kth qubit, which we denote CZ_{jk}. This has the following effect on a basis state:

$$CZ_{jk} |x\rangle = (-1)^{x^T M_{jk} x} |x\rangle , \tag{A27}$$

where M_{jk} is the $n \times n$ matrix with a 1 in position (j,k) and zeroes everywhere else. The set of all $\{M_{jk}\}$ form a basis for $n \times n$ binary matrices, hence we can convert any Q to any other Q' by an appropriately chosen sequence of CZ gates, leaving d and λ untouched. This completes the proof of statement 1.

Now to prove statement 2. From statement 1 any stabiliser state $|\phi\rangle$ is equivalent to:

$$|\mathcal{K}\rangle = \frac{1}{|\mathcal{K}|^{1/2}} \sum_{x\in\mathcal{K}} |x\rangle , \tag{A28}$$

up to some diagonal Clifford, for some $\mathcal{K}$. The strategy is to find a Clifford unitary U that commutes with $\mathcal{E}_D$, and converts the $2n$-qubit stabiliser state $|\mathcal{K}\rangle$ to some product of two n-qubit states $|\mathcal{K}'\rangle = |\mathcal{K}'_A\rangle \otimes |\mathcal{K}'_B\rangle$. Then we have:

$$\mathcal{R}[(\mathcal{E}_D \otimes \mathbb{1}_n) |\mathcal{K}\rangle\langle\mathcal{K}|] = \mathcal{R}\left[(\mathcal{E}_D \otimes \mathbb{1}_n)(|\mathcal{K}'_A\rangle\langle\mathcal{K}'_A| \otimes |\mathcal{K}'_B\rangle\langle\mathcal{K}'_B|)\right] \tag{A29}$$

$$= \mathcal{R}\left[\mathcal{E}_D(|\mathcal{K}'_A\rangle\langle\mathcal{K}'_A|) \otimes |\mathcal{K}'_B\rangle\langle\mathcal{K}'_B|\right] = \mathcal{R}\left[\mathcal{E}_D(|\mathcal{K}'_A\rangle\langle\mathcal{K}'_A|)\right], \tag{A30}$$

where the last step follows as $|\mathcal{K}'_B\rangle$ is a stabiliser state so makes no contribution to the robustness. The final state $|\mathcal{K}'\rangle$ can be factored as $|\mathcal{K}'_A\rangle \otimes |\mathcal{K}'_B\rangle$ provided its generator G' can be written in block matrix form as:

$$G' = \begin{pmatrix} G'_A & 0 \\ 0 & G'_B \end{pmatrix}, \tag{A31}$$

where G'_A and G'_B have n rows, and represent the generators for affine spaces $\mathcal{K}'_A$ and $\mathcal{K}'_B$.

We now show that we can always reach this form by a Clifford U_C comprised of a sequence of CNOTs targeted on the last n qubits. Such a sequence always commutes with $\mathcal{E}_D \otimes \mathbb{1}_n$. Suppose we have some $2n \times k$ generator G for an affine space $\mathcal{K}$ with $k = \dim(\mathcal{K})$:

$$G = \begin{pmatrix} G_A \\ G_B \end{pmatrix}, \tag{A32}$$

where G_A and G_B are each $n \times k$ submatrices. The full matrix G will have rank k, and G_A will have some rank $m \leq k$. Either G_A is already full rank ($m = k$), or it can be reduced to the following form by elementary column operations, which is equivalent to multiplication on the right

by a $k \times k$ matrix S:

$$G_A \longrightarrow G_A S = \begin{pmatrix} G'_A & 0 \end{pmatrix}, \tag{A33}$$

where G'_A is $n \times m$ (and hence full column rank), and 0 is $n \times (k-m)$. Multiplying G on the right by S, we interpret as a change in the choice of generating set:

$$G \longrightarrow GS = \begin{pmatrix} G_A S \\ G_B S \end{pmatrix} = \begin{pmatrix} G'_A & 0 \\ G''_B & G'_B \end{pmatrix}. \tag{A34}$$

Now, apply the Clifford U_C described by the matrix C in equation (A20). This transforms the generator to:

$$G' = CGS = \begin{pmatrix} \mathbb{1} & 0 \\ M & \mathbb{1} \end{pmatrix} \begin{pmatrix} G'_A & 0 \\ G''_B & G'_B \end{pmatrix} = \begin{pmatrix} G'_A & 0 \\ MG'_A + G''_B & G'_B \end{pmatrix}. \tag{A35}$$

Note that if G_A was already full rank, the change of generating set is not necessary. If we can set the bottom-left submatrix to zero, then $U_C \lvert \mathcal{K} \rangle$ can be factored as described above. This is possible if there exists a binary matrix M such that $MG'_A = G''_B$. But G'_A has full column rank m, so there exists an $m \times n$ left-inverse $G'^{-1}_{A,\text{left}}$ such that $G'^{-1}_{A,\text{left}} G'_A = \mathbb{1}$, where $\mathbb{1}$ is $m \times m$. Then we can set $M = G''_B G'^{-1}_{A,\text{left}}$, so that:

$$MG'_A = G''_B G'^{-1}_{A,\text{left}} G'_A = G''_B \mathbb{1} = G''_B. \tag{A36}$$

Then $G' = CGS$ is in the form (A31), so $U_C \lvert \mathcal{K} \rangle = \lvert \mathcal{K}'_A \rangle \otimes \lvert \mathcal{K}'_B \rangle$, as required. $\qquad \square$

Lemma A1 shows that if $\mathcal{E}_D$ is diagonal then for any $2n$-qubit stabiliser state $\lvert \phi \rangle$ we have that $\mathcal{R}((\mathcal{E}_D \otimes \mathbb{1}_n) \lvert \phi \rangle\langle \phi \rvert) = \mathcal{R}(\mathcal{E}_D(\lvert \mathcal{K} \rangle\langle \mathcal{K} \rvert))$ for some n-qubit affine space $\mathcal{K}$. This shows that the capacity can be calculated by maximising over just the representative states $\lvert \mathcal{K} \rangle$, proving Theorem A1. Table III illustrates the reduction in problem size. For example, whereas naively for a two-qubit channel we would need to calculate robustness for all $36{,}720$ four-qubit stabiliser states, using the result above we only need check one stabiliser state for each of the 7 non-trivial affine spaces. Cases up to five qubits are now tractable using this method.

n	Stabiliser states	Total affine spaces	Non-trivial affine spaces
2	60	11	7
3	1,080	51	43
4	36,720	307	291
5	2,423,520	2451	2419

TABLE III. Number of n-qubit stabiliser states compared with number of affine spaces. By trivial affine spaces we mean those comprised of a single element, which correspond to computational basis states. Diagonal CPTP channels act as the identity on such states.

4. Dimension of affine space

Here we make further observations that will help interpret numerical results from Section 8 of the main text.

Observation A1 (Dimension of affine space limits achievable robustness). *Suppose U is a diagonal unitary acting on n qubits, and suppose $\lvert \mathcal{K} \rangle$ is a stabiliser state associated with some affine space*

$\mathcal{K}$, $k = \dim(\mathcal{K})$. *Then* $\mathcal{R}(U\ket{\mathcal{K}}) = \mathcal{R}(U'\ket{\phi'})$ *where* $U'\ket{\phi'}$ *is a state on only* k *qubits, and* U' *is some* k-*qubit unitary. Therefore* $\mathcal{R}(U\ket{\mathcal{K}})$ *is upper-bounded by the maximum robustness achievable for a* k-*qubit state.*

Proof. We prove the result by showing that there is a sequence of Clifford gates that takes $U\ket{\mathcal{K}}$ to the product of a k-qubit state and an $(n-k)$-qubit stabiliser state. We know from Lemma A1 that for diagonal unitaries, all states with same affine space result in the same robustness, so it is enough to consider the state:

$$\ket{\mathcal{K}} = \frac{1}{\sqrt{|\mathcal{K}|}} \sum_{x \in \mathcal{K}} \ket{x}. \tag{A37}$$

A diagonal unitary will map this to:

$$U\ket{\mathcal{K}} = \frac{1}{\sqrt{|\mathcal{K}|}} \sum_{x \in \mathcal{K}} e^{i\theta_x} \ket{x}, \tag{A38}$$

where $\left\{e^{i\theta_x}\right\}$ will be some subset of the diagonal elements of U. The affine space $\mathcal{K}$ will have a generator matrix of rank k. As we saw in Lemma A1, a sequence of elementary row operations on the generator matrix can be realised by a sequence of CNOT gates. So we can use Clifford gates to transform any rank k generator matrix as:

$$G \longrightarrow G' = AG = \begin{pmatrix} \mathbb{1} \\ 0 \end{pmatrix}, \tag{A39}$$

where $\mathbb{1}$ is the $k \times k$ identity. Each element of $\mathcal{K}$ can be written $x = \sum_j g_j + h$, where $\sum_j g_j$ is some combination of columns of G, and h is a fixed shift vector. The transformation A corresponds to a sequence of CNOTs that we collect in a single Clifford unitary U_A, that acts on n-qubit computational basis states $\ket{x}$, where $x \in \mathcal{K}$, as follows:

$$U_A\ket{x} = \ket{y(x)} \otimes \ket{h'}, \tag{A40}$$

where h' is an $(n-k)$-length vector, and $y(x)$ is a k-length vector given by:

$$\begin{pmatrix} y(x) \\ h' \end{pmatrix} = Ax = \sum_j Ag_j + Ah. \tag{A41}$$

Note that $y(x)$ is only defined for $x \in \mathcal{K}$, and that h' is independent of x. Elements $x \in \mathbb{F}_2^n$ that are not in $\mathcal{K}$ could be mapped to a vector where the last $n-k$ bits are not h', but these never appear as terms of $U\ket{\mathcal{K}}$. Since U_A must preserve orthogonality, each $\ket{x}$, where $x \in \mathcal{K}$, maps to a distinct element of the k-qubit basis set $\{\ket{y}\}$. In fact, since y are length k and there are 2^k distinct elements, they must form the k-bit linear space $\mathcal{L}' = \mathbb{F}_2^k$. So we can write:

$$U_A U\ket{\mathcal{K}} = \frac{1}{\sqrt{|\mathcal{L}'|}} \sum_{y \in \mathcal{L}'} e^{i\theta'_y} \ket{y} \otimes \ket{h'} \tag{A42}$$

$$= (U'\ket{\mathcal{L}'}) \otimes \ket{h'}, \tag{A43}$$

where $\ket{\mathcal{L}'}$ is a k-qubit stabiliser state, and U' is the k-qubit diagonal unitary with $e^{i\theta'_{y(x)}} = e^{i\theta_x}$ as the non-zero elements. The state $\ket{h'}$ is a stabiliser state, so cannot contribute to the robustness of $U_A U\ket{\mathcal{K}}$, and therefore $\mathcal{R}(U\ket{\mathcal{L}(\mathcal{K})}) = \mathcal{R}(U_A U\ket{\mathcal{L}(\mathcal{K})}) = \mathcal{R}(U'\ket{\mathcal{L}'})$, where $U'\ket{\mathcal{L}'}$ is a k-qubit state. $\square$

Recall that in Section 8 of the main text, we found that highly structured examples of diagonal

unitaries U exist where $\mathcal{C}(U)$ is strictly larger than $\mathcal{R}(\Phi_U)$, whereas for all the random diagonal unitaries sampled, we found them to be exactly equal. We can now explain this by a concentration effect, in conjunction with Observation A1. The n-qubit random diagonal gates concentrate (with high probability) within a narrow range of values for the magic capacity, close to the maximum possible magic capacity for an n-qubit diagonal gate. If $\mathcal{R}(\Phi_U) < C(U)$ then by Theorem A1 we must have that $C(U) = \mathcal{R}(U |\mathcal{K}\rangle\langle\mathcal{K}| U^\dagger)$ for some affine space $\mathcal{K}$ of non-maximal dimension. However, $U |\mathcal{K}\rangle\langle\mathcal{K}| U^\dagger$ is Clifford equivalent to an $(n-1)$-qubit stabiliser state acted on by a diagonal unitary. Then $\mathcal{R}(U |\mathcal{K}\rangle\langle\mathcal{K}| U^\dagger)$ would be upper bounded by the maximum $C(\mathcal{E})$ for $(n-1)$-qubit diagonal unitaries. But if $C(\mathcal{E})$ is close to the maximum possible for n-qubit diagonal unitaries, then it is impossible for $U |\mathcal{K}\rangle\langle\mathcal{K}| U^\dagger$ to achieve the magic capacity.

Finally, we consider the special case of multi-control phase gates $M_{t,n}$, which we defined in the main text as:

$$M_{t,n} = \mathrm{diag}\big(\exp(i\pi/2^t), 1, 1, \ldots, 1\big), \quad t \in \mathbb{Z}. \tag{A44}$$

Note that the gate $M_{t,n}$ acts as the identity on states $|\mathcal{K}\rangle$ unless $\mathcal{K}$ contains the zero vector $0^n = (0, \ldots, 0)^T$, so if $0^n \notin \mathcal{K}$, we get $\mathcal{R}(M_{t,n} |\mathcal{K}\rangle) = 1$. But if $0^n \in \mathcal{K}$, then $\mathcal{K}$ is a linear subspace. So for this type of gate, to find all possible values of $\mathcal{R}(M_{t,n} |\mathcal{K}\rangle) > 1$ we need only consider linear subspaces. The following theorem implies that we actually only need solve one optimisation for each possible *dimension* of linear subspace rather than one for every linear subspace.

Theorem A2. *Consider the n-qubit gate $M_{t,n}$ defined by equation (A44), and let $\mathcal{L}_A$ and $\mathcal{L}_B$ be linear subspaces such that $\dim(\mathcal{L}_A) = \dim(\mathcal{L}_B) = k$. Then:*

$$\mathcal{R}(M_{t,n} |\mathcal{L}_A\rangle) = \mathcal{R}(M_{t,n} |\mathcal{L}_B\rangle). \tag{A45}$$

Proof. We largely repeat the arguments of Observation A1, for the special case where the phases are given by:

$$\theta_x = \begin{cases} \pi/2^t & \text{if } x = \vec{0} \\ 0 & \text{otherwise} \end{cases} \tag{A46}$$

Since $\dim(\mathcal{L}_A) = \dim(\mathcal{L}_B)$, their generator matrices G_A and G_B have the same rank. It follows from the arguments of Observation A1 that there exists an invertible C, corresponding to a sequence of CNOT gates, such that $G_B = CG_A$, and $|\mathcal{L}_A\rangle = U_C |\mathcal{L}_A\rangle$, where U_C is a unitary Clifford operation.

If we consider instead the state $M_{t,n} |\mathcal{L}_A\rangle$, which involves terms in the same basis vectors as $|\mathcal{L}_A\rangle$, we just need to track what happens to the phase $\exp(i\theta_0)$. Clearly, since any CNOT acts as the identity on $|0^n\rangle$, we obtain:

$$U_C M_{t,n} |\mathcal{L}_A\rangle = \frac{1}{2^{k/2}} \sum_{x \in \mathcal{L}_B} \exp(i\theta_x) |x\rangle = M_{t,n} |\mathcal{L}_B\rangle \tag{A47}$$

Since U_C is a reversible Clifford operation, by monotonicity of robustness of magic, equation (A45) follows. $\qquad\square$

From Theorem A2, then, to find $\mathcal{C}(M_{t,n})$, we only need calculate $\mathcal{R}(M_{t,n} |\mathcal{L}\rangle)$ for a single representative subspace for each possible value of $\dim(\mathcal{L})$. Recall that for n-qubit stabiliser states $|\mathcal{L}\rangle$, $k = \dim\mathcal{L}$ can take integer values from 0 to n. The states with $k = 0$ correspond to single computational basis states without superposition, so are unaffected by phase gates. That is, for n-qubit multicontrol phase gates we only have to calculate n robustnesses. Compare this to the number of optimisation problems we would need to solve without using the above observations (Table III).

We can go further. From Observation A1 we know that for a subspace with $\dim(\mathcal{L}) = k < n$, it must be the case that $M_{t_n} |\mathcal{L}\rangle$ is Clifford-equivalent to $(U' |\mathcal{L}'\rangle) \otimes |h'\rangle$ for the k-qubit state $|\mathcal{L}'\rangle$ and $(n-k)$-qubit computational basis state $|h'\rangle$, and some diagonal k-qubit unitary U'. By inspection of the phases given by equation (A46), U' can only be the k-qubit multicontrol gate $M_{t,k}$. This leads to the following statement:

Observation A2 (*n-qubit multicontrol gates*). *For any fixed t and n-qubit state $|\mathcal{L}\rangle$ where $\dim(\mathcal{L}) = k < n$, we have:*

$$\mathcal{R}(M_{t,n} |\mathcal{L}\rangle) = \mathcal{R}\left(M_{t,k} |\mathcal{L}'\rangle\right) \tag{A48}$$

where $|\mathcal{L}'\rangle$ is the k-qubit state with $\mathcal{L}' = \mathbb{F}_2^k$.

Linear subspace	Number of qubits, n			
dimension, k	2	3	4	5
1	1.414	1.414	1.414	1.414
2	**1.849**	1.849	1.849	1.849
3	-	**2.195**	2.195	2.195
4	-	-	**2.264**	**2.264**
5	-	-	-	2.195

TABLE IV. Final robustness after multicontrol-T gate applied to input stabiliser states $|\mathcal{L}\rangle$ with $k = \dim(\mathcal{L})$. In each column, the maximum robustness (i.e. the capacity) is highlighted red.

Observation A2 partially justifies our Conjecture 8.1 in Section 8 of the main text, that for fixed t, the maximum increase in robustness achievable for $M_{t,n}$, over any n, is given by $\mathcal{R}\left(M_{t,K} |+\rangle^{\otimes K}\right)$, for some finite number of qubits K. To unpack this claim further, let us consider the maximisation over input stabiliser states performed to calculate the capacity $\mathcal{C}$. In this Appendix, we have seen that for the family of gates $M_{t,n}$, we only need to calculate robustness for one representative input stabiliser state for each possible *dimension* of linear subspace; that is, for $M_{t,n}$ there are only n robustnesses to calculate. In Table IV we present the relevant values for the family of multicontrol-T gates ($t = 2$) and make two observations. First, looking across the rows of Table IV, notice that the values for fixed k are constant with n, assuming $k \leq n$. Indeed, this is a generic feature of the $M_{t,n}$ gates as formalised by Observation A2. Second, looking down the last column of Table IV, we see that up until $k = 4$, $\mathcal{R}(M_{t,n} |\mathcal{L}\rangle)$ increases with $\dim(\mathcal{L})$, but at $k = 5$ the value drops. With a little thought we can see that this is necessarily the case if $\mathcal{R}(\Phi_{M_{t,5}}) < \mathcal{C}(M_{t,5})$; we saw earlier that for diagonal gates U the Choi state robustness is equal to $\mathcal{R}(U |+\rangle^{\otimes n})$, and $|+\rangle^{\otimes n}$ is a representative state for the $k = n$ case.

Our current techniques limit us to five-qubit operations, so we are unable to confirm whether $\mathcal{R}(M_{t,n} |\mathcal{L}\rangle)$ continues to decrease with increasing $\dim(\mathcal{L})$. An intuition for why a decrease is plausible goes as follows. A stabiliser state $|\mathcal{L}\rangle$ with $\dim(\mathcal{L}) = k$ will have 2^k equally weighted terms when written in the computational basis, so will have a normalisation factor of $2^{-k/2}$. The non-stabiliser state $M_{t,n} |\mathcal{L}\rangle$ is identical to $|\mathcal{L}\rangle$ apart from the phase on the all-zero term $|0 \ldots 0\rangle$. As k becomes large, the amplitude of the term $\frac{e^{i\pi/2^t}}{2^{k/2}} |0 \ldots 0\rangle$ becomes very small, so that $M_{t,n} |\mathcal{L}\rangle$ has high fidelity with the stabiliser state $|\mathcal{L}\rangle$. We would therefore expect $M_{t,n} |\mathcal{L}\rangle$ to have a small robustness if k is large.