Stabilizer Quantum Mechanics and Magic State Distillation

Jeffrey Epstein
Advisor: Daniel Gottesman

March 13, 2015

Abstract

The subtheory of quantum computation known as stabilizer computation is reviewed along with a strategy for efficient classical simulation of stabilizer protocols. Magic state distillation, a method for the implementation of fault-tolerant universal quantum computation using fault-tolerant stabilizer protocols and access to imperfectly prepared non-stabilizer states, is also reviewed and discussed from the point of view of resource theories. Some new details about the geometric and combinatorial nature of the stabilizer polytope are presented, and a new magic monotone is defined.

Acknowledgments

I would like to thank Daniel Gottesman for his support throughout the process of preparing this essay. His patience and insight were invaluable to my understanding of the field.

Contents

1 Introduction

An arbitrary (pure) quantum state may be fully specified by listing coefficients of the elements of some basis for the Hilbert space in which it lives. It is often stated that quantum mechanics is exponential. The meaning of this claim is that when multiple systems are considered together, the cardinality of the bases of the Hilbert space describing the joint system grows as the product of the cardinalities of the bases of the Hilbert spaces corresponding to the individual systems. Therefore, the number of coefficients to be specified also grows in this way. If we consider, for example, a systems of n qubits, which individually have Hilbert spaces of dimension two (and therefore are described completely by two complex coefficients), we must list 2^n complex coefficients to specify the n-qubit state. Despite the exponential scaling of the description of arbitrary pure states, there is a subset of pure states, the simultaneous $+1$ eigenvalues of an Abelian subgroup of the Pauli group, that are efficiently (polynomially) representable [5, 11]. These are known as stabilizer states, and the set of convex combinations of such states as the stabilizer polytope. In Sec. 2 we review these states and their efficient representation. We also present original results on the geometric and combinatorial nature of the polytope.

It is generally believed that quantum computation is more powerful than classical computation. If this is indeed the case, then there can be no efficient classical simulation of arbitrary quantum protocols. However, there is a subset of protocols, generated by single-qubit phase, Hadamard, and controlled-NOT gates, as well as conditioning based on classical randomness and measurement outcomes, that is efficiently simulable. These are known as stabilizer computations [11]. Combined with the efficient representation of stabilizer states, this gives an efficiently simulable and representable subtheory of quantum computation, stabilizer computation. This is of equal power to classical compuation, and this fact is the subject of the Gottesman-Knill theorem. We review a constructive method of demonstrating this eqivalence.

In addition to the conceptual interest of examining stabilizer computation, there is also a practical motivation. Stabilizer operations may be realized fault-tolerantly due to the possibility of designing transversal gates to perform the elementary operations [6, 13]. These are implementations of gates on encoded qubits that have the property of causing at most a single error in the output register when faced with a single error in the input register. Unfortunately, the $\pi/8$ gate, addition of which extends stabilizer computation to universal quantum computation, can not be implemented transversally in schemes that allow transversal implementations of the other gates [14,18]. This means that some other way to perform the gate fault-tolerantly must be devised. One such method is the use of resource states known as magic states. These states may not be immediately prepared fault-tolerantly, but arbitrarily good approximations may be distilled from large numbers of imperfectly prepared states. This is known as magic state distillation [2]. We review one of the original distillation protocols, and present a new magic monotone, a function that tells us about the amount of resource associated with a state.

At present, there are few existing magic monotones. Of those that do exist, one, the relative entropy of magic [3], is intuitively defined in terms of the distance of a quantum state from the stabilizer polytope, but has no known analytical form, and is numerically intractable to estimate for systems much larger than a single qubit. Another, the mana [3], is easy to calculate, but is not defined for qubits. It is our hope that the characterization of the n-qubit stabilizer polytope begun in this essay will suggest new monotones that are both explicitly computable and defined on n-qubit states.

2 The Stabilizer Formalism

In this section, we introduce the Pauli operators, which form a basis for the Hermitian operators and are important objects of study in quantum information. We will show that they provide a real geometric representation of quantum state space. The Clifford group, the group of unitary permutations of the Pauli operators, is introduced. Stabilizer subspaces, those subspaces of Hilbert space that are invariant under Abelian subgroups of the Paulis, are discussed, and geometric and combinatoric analysis of the stabilizer polytope, the convex hull of the stabilizer states (one-dimensional stabilizer subspaces), is presented.

2.1 Pauli Operators

The mathematics of multi-qubit systems is described in terms of the algebra of the Pauli operators X, Y, and Z. These are pairwise anticommuting and square to identity. A faithful Hermitian operator representation of this algebra in the eigenbasis of the Z operator is:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & \text{-1} \end{pmatrix}. \tag{1}$$

The n-qubit Pauli group $\mathcal{P}_n$ is the group under matrix multiplication whose elements are tensor products of n single-qubit Pauli matrices with phases $\pm i, \pm 1$.

$$\mathcal{P}_n = \{\pm(i)\{I, X, Y, Z\}^{\otimes n}\} \tag{2}$$

It will often be convenient to discuss the Pauli operators without reference to their phases. For this purpose, we may consider the Tableau representation [11].

Definition 1. The Tableau representation of the single-qubit Pauli operators is the following mapping $P \mapsto r_P$:

$$\begin{array}{c|cccc} P & \pm(i)I & \pm(i)X & \pm(i)Z & \pm(i)Y \\ \hline r_P & (0\,0) & (1\,0) & (0\,1) & (1\,1) \end{array} \tag{3}$$

with group operation bitwise modulo two addition. Let $r_{P_1} = a \oplus b$ and $r_{P_2} = c \oplus d$ (where $\oplus$ is the direct sum, i.e. vector concatenation) be the binary vectors representing n_1- and n_2-qubit Pauli operators, respectively, with a and b n_1-dimensional and c and d n_2-dimensional. Then $r_{P_1 \otimes P_2} = a \oplus c \oplus b \oplus d$. We will often denote the vector representations of n-qubit Pauli operators as

$$r_P = \left(\begin{array}{c|c} r_P^{(X)} & r_P^{(Z)} \end{array} \right) \tag{4}$$

with $r_P^{(X)}$ and $r_P^{(Z)}$ n-dimensional binary vectors.

The group of Tableau vectors is Abelian, and is a projective representation of $\mathcal{P}_n$ and a faithful representation of $\mathcal{P}_n/\mathbb{Z}_4$. The Tableau representation provides a simple way of checking whether or not two Pauli operators commute. Define the symplectic inner product on the binary (row) vector space of dimension $2n$ as:

$$\langle u, v \rangle = u \begin{pmatrix} 0 & \mathbb{1}_n \\ \mathbb{1}_n & 0 \end{pmatrix} v^T \tag{5}$$

where arithmetic is performed modulo two. Two Pauli operators P_1 and P_2 commute if and only if $\langle r_{P_1}, r_{P_2} \rangle = 0$.

2.2 Representing Density Operators

The n-qubit Pauli operators with phase $+1$ form a basis for the set of Hermitian matrices of size 2^n. To see this, note first that the Pauli operators are orthogonal with respect to the Hilbert-Schmidt inner product:

$$(P_i, P_j)_{\text{HS}} = \text{Tr}\left[P_i^\dagger P_j\right] = 2^n \delta_{ij}. \tag{6}$$

There are 4^n such operators P_i, and all are Hermitian. Hermitian matrices of size 2^n have 4^n real parameters, so the Paulis form a basis. We may therefore represent any Hermitian operator ρ on n qubits as

$$\rho = \sum_{i=1}^{4^n} r_i P_i, \;\; r_i = \frac{1}{2^n} \text{Tr}\left[P_i \rho\right] \tag{7}$$

where the r_i are real. If ρ is a density operator, the trace is the expectation of the observable P_i, so for all P_i,

$$-2^{-n} \leq r_i \leq 2^{-n}. \tag{8}$$

with the upper bound saturated for $P_i = \mathbb{1}$. Because the coefficient of the identity is fixed, a density operator is specified completely by the coefficients of the $4^n - 1$ other Pauli operators. We denote the vector containing the unnormalized coefficients as $\vec{r}_\rho$, so that $\vec{r}_i = 2^n r_i$. It will be referred to as the generalized Bloch vector. This vector is an element of the solid unit hypercube. For one qubit, the set of valid generalized Bloch vectors is the unit ball in three dimensions, which is the well-loved Bloch sphere. Positivity imposes more complicated constraints in higher dimensions.

2.3 The Clifford Group

When faced with a set, a popular pastime is considering permutations of its elements. Because the Pauli operators may be used as a basis for density operators, a reasonable question to ask is: "Which permutations of the elements of $\mathcal{P}_n$ yield unitary quantum maps?". The answer is "members of the Clifford group". The Clifford group on n qubits is the unitary normalizer of the n-qubit Pauli group. To specify a member of the Clifford group, it is enough to specify how it acts on the single-qubit X and Z operators. In addition, any assignment of images of these operators that preserves commutation relations defines a valid element of the Clifford group. For a good discussion of the construction and properties of the Clifford group, see Ref. [16]. An important feature of the Clifford group is that all Clifford operations may be implemented using a polynomial number of controlled-NOT, Hadamard, and phase gates [6]. These gates form an almost-universal gateset, in that the addition of a single gate (often the $\pi/8$ gate) extends their reach to universal quantum computation.

2.4 Stabilizer Subspaces and States

Subgroups $S < \mathcal{P}_n$ of the n-qubit Pauli group may be used to specify subspaces of the Hilbert space of n qubits:

Definition 2. For some subgroup $S < \mathcal{P}_n$, the stabilizer subspace V_S is the subspace of the n-qubit Hilbert space such that $M|\psi\rangle = |\psi\rangle$ for all $M \in S$ and all $|\psi\rangle \in V_S$.

Note that in order for V_S to be non-trivial, it must be the case that $-I \notin S$. Because all Pauli operators square to identity and all pairs of Pauli operators either commute or anticommute, this implies several facts about stabilizer groups:

- $\pm iP \notin S$ for any $P \in \mathcal{P}_n$ with phase $+1$.
- $P \in S \rightarrow -P \notin S$.
- S is Abelian.

Because stabilizers are Abelian groups with all elements squaring to identity, any independent set of k Pauli operators with phase ± 1 (not including $-I$) generates a group S of size 2^k (elements of S correspond to the binary strings of length k). We would like to know the dimension of the stabilized subspace V_S:

Lemma 1. *Given a stabilizer S of size 2^k, the stabilizer subspace V_S has dimension $2^{-k}\dim(\mathcal{H})$, where $\dim(\mathcal{H})$ is the dimension of the n-qubit Hilbert space.*

Proof. Let the Pauli operators be $\{g_1, \ldots g_k\}$. Then the projector onto their simultaneous $+1$ eigenspace is

$$P_S = \prod_{i=1}^{k} \frac{1}{2}(\mathbb{1} + g_i) = 2^{-k} \sum_{u \in \{0,1\}^k} P_u \tag{9}$$

where $P_u = \prod_{i=1}^{k} g_i^{u_i}$. Notice that $P_S^2 = P_S$ if and only if all pairs g_i, g_j commute. Consider the elements P_u. Because the g_i are independent, $P_u = \mathbb{1}$ only for $u = 0$. Because all non-identity Pauli operators are traceless, we find

$$\mathrm{Tr}\left[P_S\right] = 2^{-k}\mathrm{Tr}\left[\mathbb{1}\right] = 2^{-k}\dim(\mathcal{H}) \tag{10}$$

□

A particular stabilizer group may be written as the group generated by many different sets of generators. A choice of k independent generators of a stabilizer S of size 2^k is called a *presentation*. In order to choose a presentation from the set of elements of S, we begin by picking any non-identity element. Having already chosen some generators, we may pick the next from any element of S that is not in the subgroup generated by the previously-selected generators. Many of the results in this section will follow from consideration of the different presentations of a given stabilizer group.

The Tableau representation of the Pauli operators gives a compact way to represent stabilizer groups or, equivalently, stabilized subspaces [11]:

Definition 3. Let S be a stabilizer group of size 2^k with generators $\{g_1, \ldots, g_k\}$. The Tableau representation of S is

$$R_S = \left(\begin{array}{c|c|c} r_{g_1}^{(X)} & r_{g_1}^{(Z)} & r_1 \\ \vdots & \vdots & \vdots \\ r_{g_k}^{(X)} & r_{g_k}^{(Z)} & r_k \end{array}\right) \tag{11}$$

where the rows of R_S are $2n+1$-dimensional binary vectors. The first $2n$ elements of each row are the Tableau representations of the generators g_i and the final element is 0 if $g_i \in S$ and 1 if $-g_i \in S$.

A particular application of the above lemma and the tableau representation is to the case $k = n$, when the stabilizer subspace is one-dimensional - a pure state. The tableau representing this state is a $n \times (2n+1)$ binary matrix, so the state is specified completely in polynomial space - this representation of a quantum state is efficient! This is in stark contrast to the exponentially large representation in terms of the amplitudes of each basis state. Of course, only a very special subset of pure states may be expressed as the subspaces stabilized by a subgroup of Pauli operators:

Definition 4. An n-qubit stabilizer state is one-dimensional stabilizer subspace.

These states have a simple representation in terms of their stabilizer groups S:

Lemma 2. *The density operator of a stabilizer state with stabilizer S has the form*

$$\rho_S = \frac{1}{2^n}\sum_{s \in S} s \tag{12}$$

Proof. The stabilizer state ρ_S is a pure state which is the simultaneous $+1$ eigenstate of all 2^n elements of S. This is equivalent to being the simultaneous $+1$ eigenstate of each of the n independent generators in some generating set of S. Let $\{g_i\}$ be such a set. Then the operator ρ_S is the projector onto the mutual $+1$ eigenspace:

$$\rho_S = \prod_{i=1}^{n} \frac{1}{2}\left(\mathbb{1} + g_i\right) = \frac{1}{2^n}\sum_{u \in \{0,1\}^n} \prod_{i=1}^{n} g_i^{u_i} = \frac{1}{2^n}\sum_{s_i \in S} s_i \tag{13}$$

□

Note that the Clifford group contains maps that send any stabilizer state to any other. Suppose we want to find an operation that takes some stabilizer state σ to another stabilizer state ρ. Because all generators of the stabilizer groups of σ and ρ commute, there is some Clifford operation C_σ that maps the single-qubit Z operators to the generators of σ and some map C_ρ that maps them to the generators of ρ. Then the map

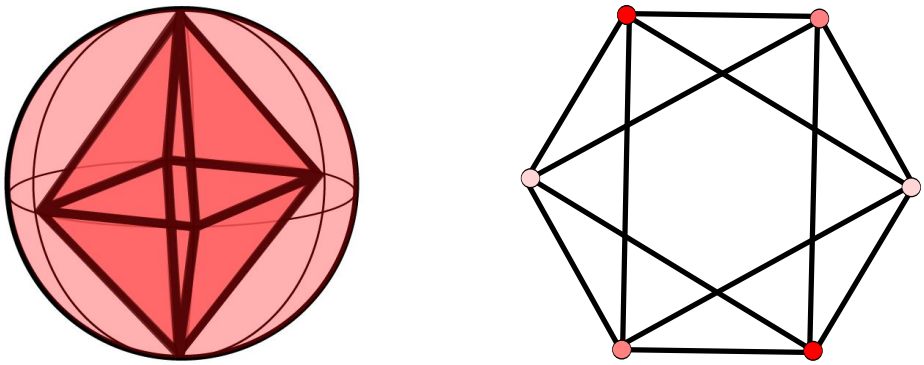


Figure 1: The single-qubit stabilizer polytope $\mathcal{P}_{\text{stab}}(1)$ has a three-dimensional representation in terms of the Bloch vectors of its elements. On the left, it is depicted as a subset of the single-qubit quantum states (the octahedron inside of the Bloch sphere). The vertices are the (pure) stabilizer states - the $+1$ eigenstates of the operators $\pm X$, $\pm Y$, and $\pm Z$. On the right, the edge graph of the polytope is shown. Vertices and edges of the graph correspond to vertices and edges of the polytope. Vertices of the same color have stabilizer groups $\{\mathbb{1}, \pm P\}$, and correspond to orthonormal bases of the qubit Hilbert space.

$C_\rho \circ C_\sigma^{-1}$ maps σ to ρ. Therefore the Clifford group acts transitively on both the Pauli group and on the set of stabilizer states.

We will be interested not only in the pure stabilizer states, but also in convex combinations of these, which may be viewed as probabilistic mixtures.

Definition 5. The n-qubit stabilizer polytope $\mathcal{P}_{\text{stab}}(n)$ is the convex hull of the set of n-qubit stabilizer states.

Computations that remain within this polytope form a special subset of quantum computations, as we will see later. One way to view the stabilizer polytope is as a solid in $4^n - 1$-dimensional Euclidean space. In this representation, the states are represented by the non-identity part of the vector $\vec{r}$ of coefficients of Pauli operators in their density operators. In this representation, the single-qubit stabilizer polytope is three-dimensional, so may be depicted as in Fig. 1.

2.5 Characterization of the Stabilizer Polytope

Now that we have defined the stabilizer polytope, we may wish to know something about its structure. If we specify stabilizers by their generalized Bloch vectors, then the n-qubit stabilizer polytope $\mathcal{P}_{\text{stab}}(n)$ is a $(4^n - 1)$-dimensional object. We may easily visualize $\mathcal{P}_{\text{stab}}(1)$. For larger n, however (even $n = 2$), we won't get very far with pictures. One way to get a handle on the structure of the polytope is by determining the spatial arrangement of the vertices. Another is figuring out which pairs of vertices define edges, i.e., finding the edge graph. In this section, we do both. We begin by counting the number of vertices of $\mathcal{P}_{\text{stab}}(n)$, i.e., the number n-qubit stabilizer states.

Theorem 3. *The number of n-qubit stabilizer states is*

$$N(n) = 2^n \prod_{k=0}^{n-1} \left(2^{n-k} + 1\right) \tag{14}$$

Proof. The proof of this fact will proceed via a probabilistic argument that follows the argument presented in Ref. [11]. Suppose that we are hoping to construct a particular stabilizer $S < P_n$ by drawing successive generators uniformly at random from the Pauli set, assigning each one a phase ± 1 as we go. We begin with all Paulis available to us except the identity, and after adding each generator to the group, we remove from the bin all Paulis generated by the elements we've already chosen as well as those that do not commute with all of the elements already chosen. In order to construct the desired stabilizer, we must at each draw pick

one of the elements of the stabilizer remaining in the bin and assign it the right phase. The probability of this is

$$\Pr_n(S) = \prod_{k=0}^{n-1} \frac{1}{2} \frac{2^n - 2^k}{2^{2n-k} - 2^k} \tag{15}$$

As we've made no assumptions about S, it must be that this probability is the same for all stabilizers. Therefore, the number of n-qubit stabilizers is simply

$$N(n) = \frac{1}{\Pr_n(S)} = 2^n \prod_{k=0}^{n-1} \left(2^{n-k} + 1\right). \tag{16}$$

□

We will now present a more convoluted procedure for counting stabilizer states. Our general strategy will be to construct stabilizer states from some "reference" stabilizer by choosing a subset of its stabilizer group to be shared with a new stabilizer state, which we will be able to build in many different ways. The challenge will be to avoid overcounting by determining the possible redundancies in descriptions of stabilizer groups. The expression we find at the end will be far less attractive than the one just derived, but will lead as a simple corollary to a characterization of the geometric structure of the stabilizer polytope. As an added bonus, preparation for this more complicated counting method requires us to take a stroll through the garden of stabilizer properties. Let us now begin this pleasant excursion.

Lemma 4. *Let $\{g_i\}$ be a set of k independent Pauli operators, let c be a binary vector of length k. There are $4 \cdot 4^n/2^k$ Pauli operators P such that $[P, g_i] = 0$ if and only if $c_i = 0$.*

Proof. Imposing commutation relations of P with the $\{g_i\}$ also imposes commutation relations of P with all elements in the group generated by these elements. We are free to choose any presentation of this group to specify commutation relations. In particular, we may always choose a presentation in which the generators are partitioned into two sets $\{h_i\}$ and $\{r_j\}$, neither of size more than n, such that $[h_i, r_j] = 0$ unless $i = j$. There is some Clifford operation $\mathcal{C}$ that maps these sets to single-qubit Z and X operators. Then we know that $\mathcal{C}(P)$ has the commutation relations with these single-qubit operators imposed by c. Then each individual relation halves the options for $\mathcal{C}(P)$ by restricting the possible values of $\mathcal{C}(P_i)$. The possible values of P are found simply by applying the inverse Clifford operation. □

Lemma 5. *Let $N_G(m)$ be the number of ordered generating sets of a stabilizer S of size 2^m.*

$$N_G(m) = \prod_{k=0}^{m-1} \left(2^m - 2^k\right) \tag{17}$$

Proof. We may choose any element of S except the identity as the first generator. In order to choose the k^{th} generator so as to enlarge the group, we must choose elements of S not generated by any of the elements already chosen. Then we have $2^m - 2^k$ options. Taking the product of the number of choices at each step yields the stated identity. □

Lemma 6. *For an arbitrary stabilizer S of size 2^n and subgroup $G < S$ of size 2^k, let $\mathcal{R}(n,k)$ be the number of subgroups $H < S$ such that $G \times H = S$.*

$$\mathcal{R}(n,k) = 2^{k(n-k)} \tag{18}$$

Proof. Let g_i be generators of G and let h_i be a set of operators that, together with G, generate S. Then the h_i generate a subgroup H such that $G \times H = S$. We may produce new subgroups H' such that $G \times H' = S$ by mutliplying the generators h_i by elements of S. Multiplying generators h_i by elements of H leaves H unchanged. We can multiply them however by elements of G in $(2^k)^{n-k}$ ways. Consider two of these ways and look at elements $g_i h_i$ and $g_i' h_i$. There is no way to make one of these elements from the group containing the other, so the groups are distinct. □

Lemma 7. *Let S_1 and S_2 be two n-qubit stabilizers, and define $S_{12} = S_1 \cap S_2$. Let $|S_{12}| = 2^k$. Then for some choice of G and H with trivial intersection such that $S_1 = S_{12} \times G$ and $S_2 = S_{12} \times H$, G and H may be written in $N_G(n-k)$ ways as the groups generated by ordered generating sets $\{g_i\}$ and $\{h_i\}$ such that $\{g_i, h_j\} = 0$ if and only if $i = j$.*

Proof. Choose arbitrary generating sets $\{g_i\}$ and $\{h_j\}$ for G and H. Define a binary matrix M such that $M_{ij} = 1$ if and only if $\{g_i, h_j\} = 1$. Adding row i to row k results in the matrix corresponding to the generator sets resulting from the substitution $g_k \to g_i g_k$, while swapping rows corresponds simply to a permutation of the generators. Therefore, we are left with valid generating sets if we permute the rows of M so that the first non-zero element of row i is not to the right of the first non-zero element of row $i+1$. Having done this, we find the smallest i such that the column j containing the first non-zero element of row i has more than one 1 in it, and add row i to all other rows with a 1 in the j^{th} column. This again leaves us with valid generating sets for G and H. Repeating this leaves us with an upper diagonal matrix encoding the commutation relations between generating sets for G and H. Now we look at the bottom-right-most corner. If there is a 1 there, add this row to all rows above that have a 1 in the final column. As above, we're left with valid generating sets. Repeat this process. If we can repeat it for all columns, we're left with a diagonal matrix, which describes generating sets of the kind posited in the statement of the fact. If this procedure fails, it is because at some point we have a 0 instead of a 1. But then there is a row that is entirely 0, corresponding to a generator g of G that commutes with all generators of H. However, because $g \in S$, g must also commute with all elements of S_{12}. Then g commutes with all elements of $S_{12} \times H = S_2$. But $g \notin H$. This is a contradiction by one of the previous facts. Therefore, we must be able to produce such generating sets. Moreover, having generated one, we can generate any other such pair of sets by performing row additions and then performing the opposite column additions (i.e., add row i to row j and then add column j to column i). There is no other way to maintain the diagonal structure of M, so there are as many such pairs of generating sets as there are generating sets of G, and fixing the presentation of G also fixes the presentation of H. □

Lemma 8. *Given a stabilizer S, any Pauli operator $P \notin S$ anticommutes with exactly 2^{n-1} elements of S.*

Proof. Let S be the stabilizer generated by $\{Z_i | i = 1, 2, \ldots, n\}$ and let $P \notin S$ be arbitrary. The elements of S are of the form $S_u = \sum_i Z_i^{u_i}$ for u an n-bit binary vector. Let $v \neq 0$ be the n-bit vector such that v_i is 0 if $P_i = I$ or Z and 1 if $P_i = X$ or Y. Then if the modulo two inner product of u and v is 0, $[S_u, P] = 0$, while if it is 1, $\{S_u, P\} = 0$. There is a bijection between elements $u \in \{0,1\}^n$ with $(u, v) = 0$ and those with $(u, v) = 1$ (simply flip the bit of u_i for i the smallest value such that $v_i = 1$). Therefore, half of the elements of S anticommute with P. This is the desired result for a particular choice of S. The Clifford group comes to the rescue, because we can map S to any other stabilizer, preserving commutation relations and subgroup inclusion. □

Theorem 9. *Let $N(n)$ be the number of n-qubit stabilizer states.*

$$N(n) = \sum_{k=0}^{n} 2^k 2^{\frac{1}{2}(n(n+3)-k(k+3))} \frac{(2^{-n};2)_n}{(2^{-k};2)_k \left(2^{-(n-k)};2\right)_{n-k}} \tag{19}$$

The symbols $(a;q)_k$ are q-Pochhammer symbols or q-shifted factorials, which for $k > 0$ are defined as:

$$(a;q)_k = \prod_{j=0}^{k-1} (1 - aq^j) \tag{20}$$

Proof. Consider an arbitrary reference stabilizer state with n-qubit stabilizer group S. How many ways can we construct another stabilizer group S' by making a series of choices about its relationship to S? To streamline the argument, we'll ignore the possible phases of the Pauli operators, and tack these choices on at the end. First we choose the size of the intersection $S \cap S'$. Because we may always find a maximum set of shared generators, the subgroup generated by which is precisely the intersection between the stabilizers, for any S' we have $|S \cap S'| = 2^k$ for some integer $0 \leq k \leq n$. Once we have chosen a size 2^k for the intersection, we choose a particular subgroup $S_{\text{int}} < S$ of that size, and construct a stabilizer S' such that $S \cap S' = S_{\text{int}}$.

Suppose that there are $\mathcal{I}(n,k)$ distinct subgroups $S_{\text{int}} < S$ of size 2^k and $\mathcal{E}(n,k)$ distinct stabilizers S' such that $S \cap S' = S_{\text{int}}$. Tacking on at the end 2^n choices about the signs of the Pauli operators (we are free to choose phases ± 1 for each of n generators) we have that the number of n-qubit stabilizers is

$$N(n) = 2^n \sum_{k=0}^{n} \mathcal{I}(n,k)\mathcal{E}(n,k) \tag{21}$$

We will now examine each of these terms in turn to find the expression given in the theorem. First, we consider the number of possible intersections (subgroups of S) of size 2^k. To choose a subgroup $S_{\text{int}} < S$ of size 2^k, we first choose a particular ordered list of n independent generators of S. Then we take the first k of these. S_{int} is the group generated by these k generators. There are $N_G(n)$ ways to do this. Of course, there is a large amount of redundancy in this procedure. First, S_{int} itself admits $N_G(k)$ ordered presentations in terms of independent generators, and the subgroup generated by the remaining $n-k$ generators of S admits $N_G(n-k)$ ordered presentations. We must also account for the fact that there are $\mathcal{R}(n,k)$ subgroups $H < S$ such that $S_{\text{int}} \times H = S$. Then we have

$$\mathcal{I}(n,k) = \frac{N_G(n)}{N_G(k)N_G(n-k)\mathcal{R}(n,k)} \tag{22}$$

Once we have chosen a particular intersection $S_{\text{int}} = S \cap S'$, we must extend it in such a way as to produce a group S' satisfying this property. In other words, we want to find groups H such that $S_{\text{int}} \times H = S$. This is equivalent to choosing $n-k$ independent generators not in S that, together with S_{int}, generate S'. Let $G < S$ be generated by the $n-k$ generators of S not chosen to be in S_{int}. Then $S_{\text{int}} \times G = S$. We know that there is a unique ordered presentation of H such that $\{h_i, g_j\} = 0$ if and only if $i = j$. Therefore we consider constructing H by choosing $n-k$ generators. They must be independent of each other and of generators of S_{int} in order to generate a complete stabilizer S'. They must also be independent of the generators of G in order that $S \cap S' = S_{\text{int}}$. Each of them then obeys n commutation relations with distinct operators, the generators of S. In addition, as we add more generators to H, the new ones must commute with those already chosen. Then after we have added j generators, we have $4^n/2^{n+j} = 2^{n-j}$ choices for the next. Having performed this procedure to extend S_{int} to a new stabilizer $S' = S_{\text{int}} \times H$, we need to take into account the multiplicity of subgroups $H < S'$ such that $S' = S_{\text{int}} \times H$. This results in the expression

$$\mathcal{E}(n,k) = \frac{\prod_{k=0}^{n-k-1} 2^{n-j}}{\mathcal{R}(n,k)} \tag{23}$$

Using these expressions for $\mathcal{I}(n,k)$ and $\mathcal{E}(n,k)$ and the expressions for $N_G(m)$ and $\mathcal{R}(n,k)$ given above, we find the expression for $N(n)$ given in the statement of the theorem. □

Equipped with an expression for the number of n-qubit stabilizer states as a sum, rather than a product, over k, we are in a position to determine the spectrum of inner products between the generalized Bloch vectors of the stabilizer states. Because these vectors are identically normalized, this tells us the distribution of distances between a given stabilizer state and all others when the stabilizer polytope is represented in (4^n-1)-dimensional real space.

Theorem 10. *Given an arbitrary vertex $\tilde{S}$ of the stabilizer polytope, the number of stabilizer states S with a particular value Ω of $\vec{r}_{\tilde{S}} \cdot \vec{r}_S$ is given by*

$$N_n(\Omega) = \begin{cases} \sum_{k=0}^{n} \left(2^k - 1\right) Q(k) & \Omega = -1 \\ Q(k) & \Omega = 2^k - 1; k = 0, \ldots, n \end{cases} \tag{24}$$

$$Q(k) = 2^{\frac{1}{2}(n(n+3)-(k(k+3))} \frac{\left(2^{-n};2\right)_n}{\left(2^{-k};2\right)_k \left(2^{-(n-k)};2\right)_{n-k}} \tag{25}$$

Proof. Once we have established a particular set of k generators to be shared by S and $\tilde{S}$, we are free to assign phases ± 1 to each of these generators of S in 2^k ways. Consider choosing a binary string u of length k and assigning phase $(-1)_i^u$ to generator g_i. Any Pauli in the intersection S_{int} may be written as $P_v = \prod_{i=1}^{k} g_i^{v_i}$

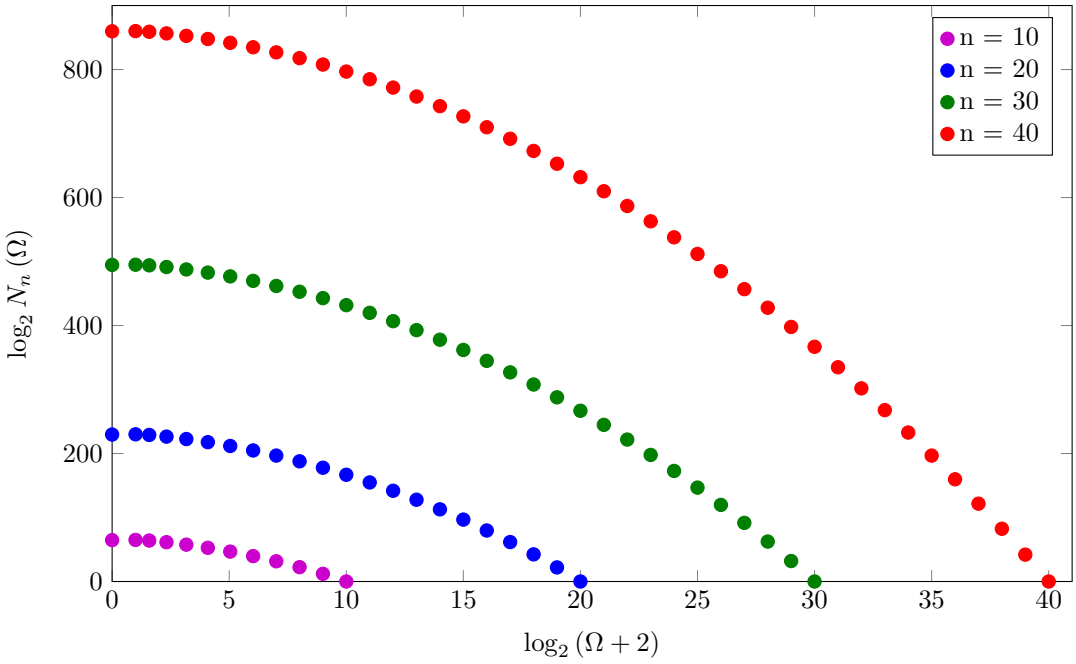


Figure 2: The number $N_n(\Omega)$ of n-qubit stabilizer states S with generalized Bloch vectors having inner product Ω with the generalized Bloch vector of some fixed n-qubit stabilizer state $\tilde{S}$ for multiple values of n.

for some other binary string v of length k. There is a sign difference between the Pauli in S and the Pauli in $\tilde{S}$ if and only if $u \cdot v = 1$. As above, this is true for exactly half of the strings v when $u \neq 0$, and the identity always has sign $+1$, so the inner product of the generalized Bloch vectors is -1 independent of k. There are $2^k - 1$ strings u for which this is the case. In the case $u = 0$, no phases are flipped, and the inner product is $2^k - 1$. □

Although this expression is horrifying, it yields a surprisingly nice form when plotted on a log-log plot as in Fig. 2. Having elucidated some of the geometric structure of the n-qubit stabilizer polytope, we turn to its combinatorial structure. In the following theorem, we characterize the edge graph of the polytope. This is the graph that has the pure stabilizer states as vertices and has an edge connecting two stabilizer states Q_1 and Q_2 if and only if no convex combination of Q_1 and Q_2 admits a decomposition in terms of other pure stabilizer states. The proof will proceed by the usual trick - mapping the stabilizers to easily manipulable stabilizers with single-qubit generators by use of the Clifford group. Then we simply examine the coefficients of each Pauli operator in the convex combination of stabilizers to determine whether distinct decompositions are possible.

Theorem 11. *The pure stabilizer states corresponding to stabilizer groups Q_1 and Q_2 form an edge of the stabilizer polytope if and only if there are Pauli operators P such that $P \in Q_1$ and $\pm P \notin Q_2$.*

Proof. A pair of stabilizers states with stabilizer groups Q_1 and Q_2 form an edge of the stabilizer polytope if and only if for all p, there is no other set of stabilizers S_j and probabilities q_j such that

$$\sum_j p_j \sum_{P \in S_j} P = p \sum_{P \in Q_1} P + (1-p) \sum_{P \in Q_2} P. \tag{26}$$

In other words, the representation of convex combinations ρ of Q_1 and Q_2 in terms of Q_1 and Q_2 is unique among representations in terms of pure stabilizers states.

For convenience of notation, for a set P of Pauli operators, let $\bar{P}$ denote the set obtained by flipping the signs of all operators in P, and let $A + B$ denote $A \cup B$ and $A - B$ denote $A \cup \bar{B}$. For pure stabilizer states with stabilizer groups Q_1 and Q_2, define the following pairwise disjoint sets: $P_+ = Q_1 \cap Q_2$, $P_- = Q_1 \cap \bar{Q}_2$,

and $P_i = Q_i \cap (Q_j)^c \cap \left(\bar{Q}_j\right)^c$ for $i \neq j$. Note that P_+ is never empty, as the identity is always an element of a stabilizer group. Then we may write $Q_1 = P_+ + P_- + P_1$ and $Q_2 = P_+ - P_- + P_2$. Then a convex combination of the two states has density operator

$$\rho = p \sum_{P \in Q_1} P + (1-p) \sum_{P \in Q_2} P = \sum_{P \in P_+} P + (2p-1) \sum_{P \in P_-} P + p \sum_{P \in P_1} P + (1-p) \sum_{P \in P_2} P. \tag{27}$$

For convenience, the states have been normalized to trace 2^n.

Following the argument made earlier in this paper, we see that we may write the two stabilizer groups in the form $Q_1 = Q_+ \times Q_- \times H$ and $Q_2 = Q_+ \times \bar{Q}_- \times G$ for some (not necessarily unique) H and G such that $H \cap Q_2 = \varnothing$ and $G \cap Q_1 = \varnothing$. We may then choose presentations

$$Q_1 = \left\langle \{k_i^{(+)}\} \cup \{k_i^{(-)}\} \cup \{h_i\} \right\rangle, \;\; Q_2 = \left\langle \{k_i^{(+)}\} \cup \{-k_i^{(-)}\} \cup \{g_i\} \right\rangle \tag{28}$$

such that all the $k_i^{(\pm)}$, h_i, and g_i are independent, the $k_i^{(\pm)}$ commute with the h_i and g_i, and $[h_i, g_j] = 0$ unless $i = j$. Then there is a Clifford group operation $\mathcal{C}$ such that

$$\mathcal{C}(Q_1) = \langle \{Z_i\} \cup \{Z_j\} \cup \{Z_k\} \rangle, \;\; \mathcal{C}(Q_1) = \langle \{Z_i\} \cup \{-Z_j\} \cup \{X_k\} \rangle \tag{29}$$

where $i \in \{1, \ldots, I\}$, $j \in \{I+1, \ldots, I+J\}$, and $k \in \{I+J+1, \ldots, n\}$ for some $I + J \leq n$. If we can find some other stabilizers S_j and probabilities q_j such that

$$\sum_j p_j \sum_{P \in S_j} P = p \sum_{P \in \mathcal{C}(Q_1)} P + (1-p) \sum_{P \in \mathcal{C}(Q_2)} P \tag{30}$$

then we have

$$\sum_j p_j \sum_{P \in \mathcal{C}^{-1}(S_j)} P = p \sum_{P \in Q_1} P + (1-p) \sum_{P \in Q_2} P. \tag{31}$$

Therefore, we need only consider pairs of stabilizer states with generating sets composed of single-qubit X and Z operators, a much more concrete and manipulable problem than the general case. This decomposition allows us to break the problem down into three cases.

Case 1: P_+ is non-empty, $P_- = \varnothing$ and $P_1, P_2 = \varnothing$. This is simply the case in which $Q_1 = Q_2 = P_+$. Then ρ is a pure state, so may not admit any other decomposition in terms of pure states. Therefore, $Q_1 = Q_2$ defines a vertex of the stabilizer polytope.

Case 2: P_+, P_-, P_1, and P_2 are all non-empty. How can the elements of Q_1 and Q_2 be distributed among the S_j? Clearly we must have $P_+ < S_j$ for all j, because the coefficient of P_+ must be one. Now consider Z_k and X_k. These anticommute, so may not both be elements of the same S_j, and their coefficients in the convex combination of Q_1 and Q_2 are p and $1-p$, respectively. Therefore, we have:

$$p = \sum_{j | Z_k \in S_j} q_j, \;\; 1 - p = \sum_{j | X_k \in S_j} q_j. \tag{32}$$

Then the S_j are partitioned into two sets, one with X_k and the other with Z_k. Now suppose that there is some j such that $Z_k, X_{k'} \in S_j$ for $k \neq k'$. The coefficient of $Z_k Z_{k'}$ in the convex combination of the S_j is

$$\sum_{j | Z_k, Z_{k'} \in S_j} q_j < \sum_{j | Z_k \in S_j} q_j = p. \tag{33}$$

This is a contradiction. Then the S_j are partitioned into two sets, one with all single-qubit Z operators on the last $n - I - J$ qubits and the other with all single-qubit X operators on the last $n - I - J$ qubits.

Now we move on to the last set of generators, the $\pm Z_j$. Suppose that for some j, $-Z_j, Z_k \in S_j$. Then the coefficient of $Z_j Z_k$ in the convex combination is

$$\sum_{j | Z_j, Z_k \in S_j} q_j - \sum_{j | -Z_j, Z_k \in S_j} q_j < \sum_{j | Z_k \in S_j} q_j = p. \tag{34}$$

This is a contradiction. Therefore, for all j, either $S_j = Q_1$ or $S_j = Q_2$. Then the decomposition in terms of the S_j is not distinct from that in terms of the Q_i, so the Q_1 and Q_2 define an edge of the polytope.

Case 3: P_+ and P_- are non-empty, $P_1, P_2 = \varnothing$. Then $\mathcal{C}(Q_1)$ and $\mathcal{C}(Q_2)$ are both computational basis states:

$$\mathcal{C}(Q_1) = |0\ldots0\rangle|0\ldots0\rangle := |0_L\rangle \tag{35}$$
$$\mathcal{C}(Q_2) = |0\ldots0\rangle|1\ldots1\rangle := |1_L\rangle. \tag{36}$$

Define also $|+_L\rangle = \frac{1}{\sqrt{2}}(|0_L\rangle + |1_L\rangle)$ and $|-_L\rangle = \frac{1}{\sqrt{2}}(|0_L\rangle - |1_L\rangle)$. We can use these as codewords corresponding in the obvious way to the single-qubit states $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|-\rangle$. It is easily verifiable that

$$p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = (1-p)|+\rangle\langle +| + (1-p)|-\rangle\langle -| + (2p-1)|0\rangle\langle 0|. \tag{37}$$

For $p \geq \frac{1}{2}$ (for $p < \frac{1}{2}$ we may simply swap the roles of $|0\rangle$ and $|1\rangle$), the right hand side is a convex combination of (single logical-qubit) stabilizer states, so applying the decoding operation, we find convex combination of three stabilizer states that is equal to the combination $p\mathcal{C}(Q_1) + (1-p)\mathcal{C}(Q_2)$. Therefore, Q_1 and Q_2 do not define an edge of the polytope. □

Note that for $Q_1 \neq Q_2$ it is *not* the case that the representation of a state as a convex combination of Q_1 and Q_2 is unique among representations in terms of convex combinations of *arbitrary* pure states. Indeed, any mixed state admits infinitely many decompositions in terms of pure states. Also note that stabilizer bases for the n-qubit Hilbert space correspond to subsets $B \subset V$ of the vertices of the stabilizer polytope such that all elements of B are not adjacent in the edge graph of the polytope. The edge graph of the $n = 1$ stabilizer polytope is depicted in Fig. 1. Already for $n = 2$, the graph is useless from a visualization point of view (it is almost the complete graph on sixty vertices).

3 Stabilizer Protocols and Efficient Classical Simulation

It is often interesting to define subtheories of quantum mechanics or computation by restricting the allowed operations to a subset of those allowed by the complete theory. One example of this type of restriction that leads to interesting properties is stabilizer computation. To define this, we follow [2, 3].

Definition 6. A stabilizer protocol is a protocol consisting of only the following operations:

- Preparation of qubits in the state $|0\rangle$
- Clifford operations
- Measurement in the Z basis on the final qubit
- Partial trace
- Any of these operations conditioned on the outcomes of measurements or on classical randomness

Definitions of stabilizer computation differ slightly from author to author. For instance, some allow for preparation of any stabilizer state or measurement of any Pauli operator. However, due to the transitive action of the Clifford group on the Paulis and on the stabilizer states, these operations are related to the ones allowed here by a Clifford operation. This means, first of all, that the protocol is itself a stabilizer protocol, but also, because a Clifford gate may be implemented with a polynomial number of elementary gates, that the models have equivalent power in terms of the comutational complexity of any algorithm.

Stabilizer protocols are not universal for quantum computation. For instance, any pure state other than the stabilizer states is inaccessible. Clifford operations also do not generate the full unitary group. However, they are still capable of performing many interesting tasks usually associated with quantum behavior such as quantum teleportation, superdense coding, and quantum error correction [6]. Note that, because we may simulate classical computation efficiently by a stabilizer protocol (simply by using only the qubit basis states with no superpositions), stabilizer protocols are at least as powerful as classical computation.

In fact, stabilizer protocols have the same power as classical computation, because they may be efficiently simulated by a classical computer. To demonstrate this, it is necessary first to decide what exactly we mean by simulation of a quantum computer. In [17], a distinction is drawn between strong and weak simulation. A strong simulation of a quantum computation is a classical algorithm that produces a probability distribution for the possible outcomes. A weak simulation is a classical algorithm that samples from this distribution. In discussing the power of subtheories of quantum computation, it is reasonable to consider weak simulation, because the output is the same as the output of the quantum computer under consideration: a random number sampled from a distribution.

The Gottesman-Knill theorem concerns the simulability of stabilizer computation. Typical proofs of the theorem give explicit methods of simulation. Both weak and strong efficient classical simulation are attainable. The following statement of the theorem is quoted from Ref. [6].

Theorem 12. *"Suppose a quantum computation is performed which involves only the following elements: state preparation in the computational basis, Hadamard gates, phase gates, controlled-NOT gates, Pauli gates, and measurements of observables in the Pauli group (which includes measurement in the computational basis as a special case), together with the possibility of classical control conditioned on the outcome of such measurements. Such a computation may be efficiently simulated on a classical computer."*

To demonstrate the possibility of efficient classical weak simulation of stabilizer protocols, i.e., to prove the theorem, an explicit construction is presented. This construction was first described in [5] and was expanded in [11].

The strategy for simulating a stabilizer protocol will be to keep track of the tableau representation of the state of the system. Because the system is initialized in a stabilizer state and only subjected (possible governed by a classical probability distribution) to operations that map stabilizer states to other stabilizer states, the state will always admit such a representation. As discussed earlier, this representation is efficient. Here we will give efficient methods for simulating the possible operations. Probabilistic aspects of the protocols will simply be imported directly into the classical protocol. Together, these will prove that stabilizer computation is no more powerful than classical computation. In part for simplicity and convenience, and in part because it might be desirable actually to implement these simulation techniques, I will present these methods in as close to a pseudocode manner as possible.

Preparation of Qubits in the State $|0\rangle$ If we begin with an n-qubit stabilizer state with stabilizer $S = \langle s_i \rangle$ and tensor in an additional single qubit (which we will assign index $n+1$) in the state $|0\rangle$, the resulting state is stabilized by $S \otimes \mathbb{1} \times \mathbb{1}_n \otimes Z$. Therefore, a generating set for the group is $\{s_i\} \cup Z_{n+1}$. We may therefore represent this operation by updating the tableau as follows:

$$R = \left[\begin{array}{c|c|c} R_X & R_Z & R_P \end{array}\right] \longrightarrow R' = \left[\begin{array}{cc|cc|c} R_X & 0 & R_Z & 0 & R_P \\ 0 & 0 & 0 & 1 & 0 \end{array}\right]. \tag{38}$$

Clifford Operations The generators of the Clifford gates are the Hadamard, phase, and controlled NOT gates. These have the following actions on the generators of the Pauli group:

$$H : \left\{\begin{array}{l} X \mapsto Z \\ Z \mapsto X \end{array}\right. \qquad S : \left\{\begin{array}{l} X \mapsto Y \\ Z \mapsto Z \end{array}\right. \qquad CNOT : \left\{\begin{array}{l} XI \mapsto XX \\ IX \mapsto IX \\ ZI \mapsto ZI \\ IZ \mapsto ZZ \end{array}\right. . \tag{39}$$

These actions lead to rules for updating the stabilizer tableau. `ADDTO`(i,j) calls for column i to be added to column j. `TIMES`(i,j) returns the bitwise product of columns i and j, while `PLUS` returns the bitwise sum. `SWAP`(i,j) switches columns i and j. All arithmetic is performed modulo two.

```
function [tableau] = Clifford(tableau,gate,i,j)

[n,-] = size(tableau);

if gate = H
    addto(times(i,i+n),2n+1);
    swap(i,i+n);
elseif gate = S
    addto(times(i,i+n),2n+1);
    addto(i,i+n);
elseif gate = CNOT
    addto(times(times(i,j+n),plus(j,i+n,1)),2n+1);
    addto(i,j);
    addto(n+j,n+i);
end
```

The first step of each operation is an update of the phase column. The only time this need be done is for Y in both phase and Hadamard gates, and for YY and XZ in CNOT. Constructing truth tables for the terms being added to the phase column demonstrates that these are precisely the conditions under which a bit flip is performed.

Measurements in the Z basis on the Final Qubit There are three scenarios we must consider. In the case that $Z \in S$, a measurement returns the value $+1$ and leaves the state unchanged. If $-Z \in S$, a measurement returns -1 and leaves the state unchanged. If $\pm Z \notin S$ the probability of measuring $+1$ is one half. We simulate sampling from this distribution by flipping a fair coin. In this case, however, we must also update the stabilizer. We know that $Z \notin S$ implies that Z commutes with exactly half of the members of S. The eigenvalue of these operators should remain unchanged by measurement of Z, so we leave them in S. These correspond to $n-1$ generators. Then we simply remove the generator that doesn't commute with Z and replace it by Z. This results in the stabilizer of the post-measurement state. We represent this in the tableau by leaving alone any generator that has a zero in the n^{th} column (corresponding to Z_n or I_n). We replace the first row (call it r) that has a one in the n^{th} column with the row corresponding to Z_n. Then we multiply every other row Pauli P with X_n by the Pauli corresponding to the first row. This corresponds by adding r to every other row m that has a one in the n^{th} column. Because rm is in S, the result is a tableau with $n-1$ generators that are in S and commute with Z_n. This then is the generating set for the stabilizer of the post-measurement state.

4 Magic - Universal Quantum Computation and Resource Theory

In the last section, we defined a particular subtheory of quantum computation, stabilizer computation, and demonstrated its equivalence to classical computation from a mutual efficient simulability point of view. In this section, we will consider those states and operations that are not accessible from the stabilizer framework. States that may not be prepared by any stabilizer protocol are known as *magic states* [3]. We will see that these states can provide a resource, *magic*, that allows stabilizer computation to be extended to fault-tolerant universal quantum computation (UQC) via magic state distillation, first introduced in Ref. [2]. We will also discuss *magic monotones*, functions that measure the amount of magic possessed by quantum states. A new monotone will be introduced. We begin with a brief discussion of error correction and fault-tolerance.

4.1 Error Correction

This discussion of QECCs follows very loosely the much more detailed account in Ref. [6]. Suppose that the letters x of some alphabet correspond to the states $|\psi_x\rangle$ of some Hilbert space $\mathcal{H}_{\text{clear}}$ that also describes a physical system. We would like to be able to send a message by sending a sequence of these systems prepared in the appropriate states. Unfortunately, between the preparation of the message and its receipt, errors may

occur due to environmental noise. Suppose that these errors are described by quantum operations from some set $\mathcal{E}$. We would like to be able to detect and correct these errors in some way. In classical communication, a common way to do this is using redundancy. We can simply send some large number of copies of each letter, and then apply a majority-voting procedure to decode. In the quantum case, this is note possible due to the no-cloning theorem [6]. However, we can still use the notion of redundancy.

Consider an auxiliary system described by the Hilbert space $\mathcal{H}_{\text{aux}}$ and define $\mathcal{H}_{\text{crypt}} = \mathcal{H}_{\text{clear}} \otimes \mathcal{H}_{\text{aux}}$. Sending systems described by $\mathcal{H}_{\text{crpyt}}$ trivially introduces redundancy in our message, because there are $\dim(\mathcal{H}_{\text{aux}})$ orthogonal states in $\mathcal{H}_{\text{crypt}}$ for each state $|\psi_x\rangle \in \mathcal{H}_{\text{clear}}$. In practice, however, this is unlikely to be a useful encoding. For example, the errors $\mathcal{E}$ may be local in some sense (for instance, one- or two-qubit operators if the systems are systems of qubits). Therefore, elements of $\mathcal{E}$ may map $|\psi_x\rangle \otimes |\psi\rangle$ to $|\psi_{x'}\rangle \otimes |\psi\rangle$ for some $x' \neq x$. Then there is no way to distinguish between the cases in which an error occurred and the cases in which x' was the intended letter.

To deal with this problem, we would like somehow to spread the information throughout the entirety of the physical system corresponding to $\mathcal{H}_{\text{crypt}}$. To do so, we apply an encoding operation described by a unitary operator $\mathcal{U}_{\text{en}}$ on $\mathcal{H}_{\text{crypt}}$. This has the effect of mapping the elements of $\mathcal{H}_{\text{plain}}$ (the letters) onto a possibly non-separable (with respect to the plain/aux partition) subspace of $\mathcal{H}_{\text{crypt}}$. This subspace is known as the codespace or QECC. If $\mathcal{U}_{\text{en}}$ was chosen appropriately given $\mathcal{E}$, then errors will map states in the $QECC$ subspace to states in other subspaces of $\mathcal{H}_{\text{crypt}}$, with the same error sending all states to the same subspace. These subspaces are distinguished by their eigenvalues under some set of operators. Measuring these operators, we determine the *syndrome* S of the error (the collection of eigenvalues), and can apply a correction operator conditioned on the syndrome. This operator maps the subspace $\mathcal{H}_S$ corresponding to the measured syndrome to the QECC, so we can simply apply $\mathcal{U}_{\text{en}}^{-1}$ to recover the cleartext of our message. Note that it is by no means clear from this discussion how to go about *finding* encoding operations and syndrome measurements with the necessary properties.

The method of error detection described here yields a result that might be surprising: it suffices to be able to detect a finite number of errors, despite the fact that an (uncountably) infinite number of errors may occur. This convenient feature is due to the measurement step of the protocol, which projects the noisy state onto one of a finite number of error subspaces.

4.2 Fault-Tolerant UQC via Magic State Distillation

Given the large array of possible sources of error in any physical implementation of quantum computation or quantum information processing, it is important to be able to implement these schemes fault-tolerantly. One way in which this may be achieved is via concatenated QECCs. Concatenation of QECCs is the procedure of applying one encoding circuit to the output of another, creating "layers" of encoding. The threshold theorem [13] demonstrates that concatenated codes may be used to achieve arbitrarily low logical error rates provided the physical error rate is below a certain threshold, and provided that the encoded gates not propagate single-qubit errors to too many other qubits. In the original work of Aharonov and Ben-Or, this latter constraint is formalized via the notion of spread, the largest number of qubits in the output register of a block affected by a single error in that block. In [6], the presentation of QECCs is in terms of encoded gate implementations with spread 1, which are known as transversal implementations. Many QECCs admit transversal implementations of some gates, for example the Steane code and Hadamard, phase, and CNOT gates. Unfortunately, it has been shown that there is no QECC that admits a universal set of transversal gates [14, 18]. This motivates the design of schemes for universal quantum computation that do not rely on transversal universal gate sets.

Subtheories of quantum computation are conveniently specified by the set of allowed elementary operations: preparations; transformations; and measurements. Often, forbidden elementary operations of one type may be performed using available elementary operations of another type. We have already seen this in the computational equivalence of the multiple definitions of stabilizer computation. This feature also allows restricted models of computation to be extended by access to additional resources. A well-known example is that of LOCC, the set of computations allowing local operations and classical communication. This subtheory is derived from universal quantum computation by the removal of entangling gates between two subsystems. If a Bell pair is shared between these subsystems, however, a controlled-NOT gate may be implemented between them (Fig. 3).

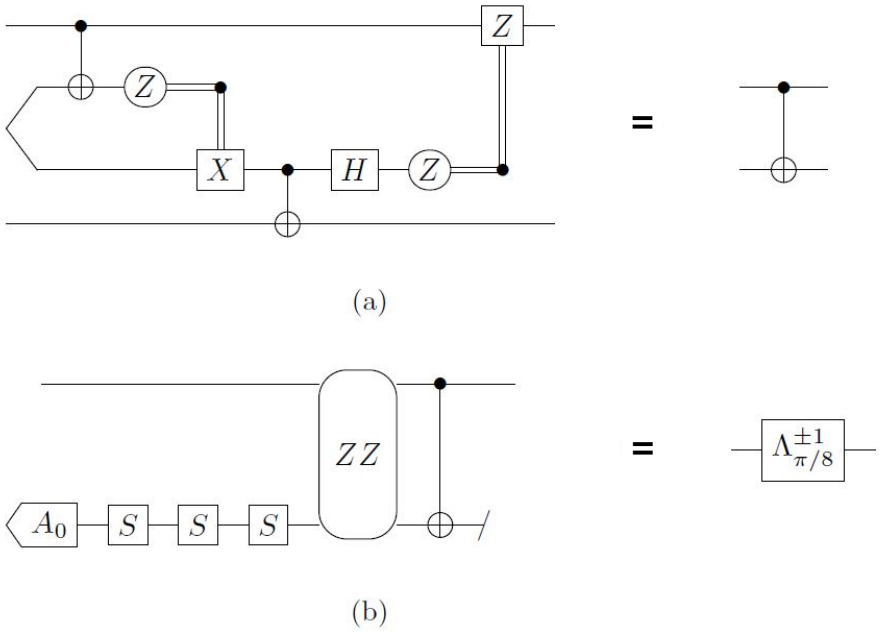


Figure 3: Restricted gatesets can be used to perform universal quantum computation if access to certain resource states is allowed. (a) Access to a single Bell pair $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ allows a nonlocal CNOT gate to be performed using only local operations and classical control (LOCC) [15]. Note that there is a well-defined sense of locality in the circuit on the left, which is specified by the choice of forbidden CNOT gate. The circles are measurements in the basis indicated inside. (b) Access to a single magic state allows the $\pi/8$ gate or its inverse to be performed with equal probability using only operations and measurements available in stabilizer protocols. The rounded box indicates measurement in the specified basis and the left-pointing box indicates input of a qubit in the magic state $|A_0\rangle$.

Similarly, a $\pi/8$ gate may be performed using only stabilizer operations given access to a single pure state of a particular type (Fig. 3). Because the ability to perform $\pi/8$ gates as well as Clifford gates yields universal quantum computation (UQC), access to (an unlimited number of copies of) in addition to the ability to perform stabilizer protocols allows UQC. However, such states lie outside of the stabilizer polytope. Then all that we've done is traded the need for non-stabilizer operations for the need for non-stabilizer states. Because we have assumed that only stabilizer operations may be performed fault-tolerantly, we need some way to get copies of this resource state via a stabilizer protocol from imperfectly prepared approximations.

One such strategy for universal fault-tolerant quantum computation is magic state distillation, introduced in [2]. This procedure works by starting with a large number of qubits in a state close enough (in a sense that will be made precise later) to the desired state, and then performing a particular stabilizer protocol eventually to produce a single-qubit state with arbitrarily high fidelity with the desired pure resource state. One of the methods from [2] is presented in detail in the next section.

4.3 A Particular Distillation Protocol

The original protocol for magic state distillation makes use of a particular type of QECC known as CSS codes. These are really nothing more than partitions of the Hilbert space of a quantum system into simultaneous eigenspaces of a set of operators. In order to understand how the distillation procedure works, it will be necessary first to review how measurements work in these codes.

Definition 7. The code $\mathrm{CSS}(A, \mathcal{L}_A; B, \mathcal{L}_B)$ is the following decomposition of the n-qubit Hilbert space:

$$\left(\mathbb{C}^2\right)^{\otimes n} = \bigoplus_{\mu \in \mathcal{L}_A^*} \bigoplus_{\eta \in \mathcal{L}_B^*} \mathcal{H}(\mu, \eta) \tag{40}$$

where A and B are anti-commuting Hermitian operators that square to identity and $\mathcal{L}_A$ and $\mathcal{L}_B$ are orthogonal binary vector spaces. The subspaces $\mathcal{H}(\mu, \eta)$ are defined by

$$A(u)B(v)\mathcal{H}(\mu, \eta) = (-1)^{\mu(u)+\eta(v)}\mathcal{H}(\mu, \eta) \quad \forall (u, v) \in \mathcal{L}_A \times \mathcal{L}_B \tag{41}$$

Note that $A(u)$ and $B(v)$ commute by the orthogonality of the subspaces $\mathcal{L}_A$ and $\mathcal{L}_B$. The dual vectors μ and η are known as the A and B syndromes, respectively. They are simply the eigenvalues of the operators $A(\mathcal{L}_A)$ and $B(\mathcal{L}_B)$.

In [2], a particular CSS code is considered. Let $A = \frac{1}{\sqrt{2}}(X + Y)$. Let f be a function of four Boolean variables and let $[f] \in \{0, 1\}^{15}$ be a vector containing the values of f on all values of the four variables except 0000. Let $\mathcal{L}_1$ be the linear subspace spanned by the indicator functions $[x_i]$ and let $\mathcal{L}_2$ be the linear subspace spanned by the indicator functions $[x_i]$ and $[x_i x_j]$. The subspaces have cardinalities, respectively, 2^4 and 2^{10}, so the pair $\mathcal{L}_1 \times \mathcal{L}_2$ has cardinality 2^{14}. Then each subspace $\mathcal{H}(\mu, \eta)$ of the Hilbert space $(\mathbb{C}^2)^{\otimes n}$ has dimension 2^{n-14}. We will consider the case $n = 15$, which corresponds to the encoding of a single logical qubit in fifteen physical qubits. The following lemmas about this code are proven in [2]

Lemma 13. *Consider two* CSS *codes,* $\mathrm{CSS}_{\mathrm{ZA}}(\mathrm{Z}, \mathcal{L}_2; \mathrm{A}, \mathcal{L}_1)$ *and* $\mathrm{CSS}_{\mathrm{ZX}}(\mathrm{Z}, \mathcal{L}_2; \mathrm{X}, \mathcal{L}_1)$ *associated, respectively, with the decompositions*

$$\left(\mathbb{C}^2\right)^{\otimes n} = \bigoplus_{\mu \in \mathcal{L}_2^*} \bigoplus_{\eta \in \mathcal{L}_1^*} \mathcal{H}_{ZA}(\mu, \eta) = \bigoplus_{\mu \in \mathcal{L}_2^*} \bigoplus_{\eta \in \mathcal{L}_1^*} \mathcal{H}_{ZX}(\mu, \eta) \tag{42}$$

For any $\eta \in \mathcal{L}_1^*$, $\mathcal{H}_{ZA}(0, \eta) = \mathcal{H}_{ZX}(0, \eta)$ *and for any* $\mu \in \mathcal{L}_2^*$ *there is some* $w \in \{0, 1\}^{15}$ *such that for any* $\eta \in \mathcal{L}_1^*$, $\mathcal{H}_{ZA}(\mu, \eta) = A(w)\mathcal{H}_{ZX}(0, \eta)$.

Lemma 14. *The following hold for the vector spaces* $\mathcal{L}_1$ *and* $\mathcal{L}_2$*:*

- $\mathcal{L}_{1/2}^{\perp} = \mathcal{L}_{2/1} \oplus [1]$
- *For all* $u \in \mathcal{L}_1$, $|u| \equiv 0 \mod 8$.

We are now in a position to be able to describe the distillation procedure:

1. Prepare 15 copies of the approximate magic state ρ. Initial error probability is defined as $\epsilon = \langle A_1|\rho|A_1\rangle$.

2. Dephase each copy of ρ by applying A with probability one half, i.e., $\rho \to \frac{1}{2}\left(\rho + A\rho A^\dagger\right)$, ensuring that the density operators are diagonal in the eigenbasis $\{|A_0\rangle, |A_1\rangle\}$ of the operator A.

3. Measure the operators $Z(\mathcal{L}_2)$ to determine the Z-syndrome $\mu \in \mathcal{L}_2^*$.

4. Find $w \in \{0,1\}^{15}$ such that for any $\eta \in \mathcal{L}_1^*$, $\mathcal{H}_{ZA}(\mu, \eta) = A(w)\mathcal{H}_{ZX}(0, \eta)$ and apply $A(w)^\dagger$.

5. Measure the operators $X(\mathcal{L}_1)$ to find η. Declare failure if $\eta \neq 0$.

6. Apply the Clifford operation that maps $X^{\otimes 15}$, $Y^{\otimes 15}$, and $-Z^{\otimes 15}$ to the single-qubit X, Y, and Z operators on the first qubit.

7. Discard qubits 2-14.

The procedure works as follows (all calculations have been deferred to Appendix A):

In the $\{|A_0\rangle, |A_1\rangle\}$ basis, application of the dephasing channel effects the following transformation on each single-qubit state:

$$\rho = \begin{pmatrix} 1-\epsilon & \rho_{01} \\ \rho_{10} & \epsilon \end{pmatrix} \to \begin{pmatrix} 1-\epsilon & 0 \\ 0 & \epsilon \end{pmatrix} = (1-\epsilon)|A_0\rangle\langle A_0| + \epsilon|A_1\rangle\langle A_1| := \rho_{\mathrm{d}} \tag{43}$$

Then the total fifteen-qubit state is given by

$$\rho_{\mathrm{in}} = \rho_{\mathrm{d}}^{\otimes 15} = \sum_{u \in \{0,1\}^{15}} \epsilon^{|u|}(1-\epsilon)^{15-|u|}|A_u\rangle\langle A_u| \tag{44}$$

where $|u|$ denotes the Hamming weight of a binary vector u (the number of ones) and $|A_u\rangle = A_{u_1} \otimes A_{u_2} \otimes \ldots \otimes A_{u_{15}}$. Measurement of the Z-syndrome, application of the correction operator $A^\dagger(w)$, and post-selected measurement of the X-syndrome has the effect of projection onto the subspace $\mathcal{H}_{ZA}(0,0)$. This subspace is spanned by the orthonormal basis states

$$\left|A_{0/1}^L\right\rangle = |\mathcal{L}_2|^{-\frac{1}{2}} \sum_{v \in \mathcal{L}_2} \left|A_{v+[0/1]}\right\rangle \tag{45}$$

The (unnormalized) fifteen-qubit state after this projection is

$$\rho_{\mathrm{s}} = |\mathcal{L}_2|^{-1} \sum_{u \in \mathcal{L}_2} \epsilon^{|u|}(1-\epsilon)^{15-|u|}\left|A_0^L\right\rangle\left\langle A_0^L\right| + |\mathcal{L}_2|^{-1} \sum_{u \in \mathcal{L}_2} \epsilon^{15-|u|}(1-\epsilon)^{|u|}\left|A_1^L\right\rangle\left\langle A_1^L\right| \tag{46}$$

The probability of success for the subroutine is then

$$p_{\mathrm{s}} = |\mathcal{L}_2| \operatorname{Tr}[\rho_{\mathrm{s}}] = \sum_{u \in \mathcal{L}_1^\perp} \epsilon^{15-|u|}(1-\epsilon)^{|u|} \tag{47}$$

The factor of $|\mathcal{L}_2|$ comes from the fact that any measured value of μ may lead to success. Now we may apply the decoding operation, which maps the state of the logical qubit to the state of a single physical qubit, i.e., $C : \alpha\left|A_0^L\right\rangle\left\langle A_0^L\right| + \beta\left|A_1^L\right\rangle\left\langle A_1^L\right| \mapsto \alpha|A_0\rangle\langle A_0| + \beta|A_1\rangle\langle A_1|$. This yields the (normalized) output state

$$\rho_{\mathrm{out}} = \frac{|\mathcal{L}_2|^{-1} \sum_{u \in \mathcal{L}_2} \epsilon^{|u|}(1-\epsilon)^{15-|u|}|A_0\rangle\langle A_0| + |\mathcal{L}_2|^{-1} \sum_{u \in \mathcal{L}_2} \epsilon^{15-|u|}(1-\epsilon)^{|u|}|A_1\rangle\langle A_1|}{\operatorname{Tr}[\rho_{\mathrm{s}}]} \tag{48}$$

From this expression, we may read off the post-subroutine error probability

$$\epsilon_{\mathrm{out}} = \frac{\sum_{v \in \mathcal{L}_2} \epsilon^{15-|v|}(1-\epsilon)^{|v|}}{\sum_{v \in \mathcal{L}_1^\perp} \epsilon^{15-|v|}(1-\epsilon)^{|v|}} \tag{49}$$

To distill a good approximation of the state $|A_0\rangle$, we begin with a large number N of poorer approximations $\rho^{(0)}$ with error probability $\epsilon^{(0)}$, partition them into groups of fifteen, and apply the distillation subroutine. In

the cases in which the subroutine succeeds, the output state is some $\rho^{(1)}$ with error probability $\epsilon^{(1)}$. Again, we partition the (at most $N/15$) copies of $\rho^{(1)}$ into groups of fifteen and apply the distillation subroutine to each group. This process continues as long as there are enough copies of the state, eventually yielding a single state $\rho^{(n)}$ with error probability $\epsilon^{(n)}$ after n iterations. However, in general $\epsilon^{(n)}$ is not smaller than $\epsilon^{(0)}$. In Fig. 4a the output error of the distillation subroutine is plotted against input error. The threshold error probability is found numerically to be $\epsilon_{\text{th}} \approx .14$. Also relevant is the probability of subroutine success, plotted in Fig. 4b. Because success probability tends to one as input error probability goes to zero, distillation succeeds in the case $\rho^{(0)} < \rho_{\text{th}}$, in the sense that in the limit of large N, the output state n is an arbitrarily good approximation of $|A_0\rangle$.

Note that what we've referred to as a physical qubit throughout this section will not, in fact, be a physical qubit in any implementation of magic state distillation. Rather, it will be an encoded qubit. Because we know how to perform the stabilizer operations that make up the distillation protocol fault-tolerantly on encoded qubits, we may forget about this fact. Also note that this discussion has proceeded from the assumption that we may prepare as many identical copies as we like of some resource state $\rho^{(0)}$. This of course will not be the case, due to the fact that $\rho^{(0)}$ must be a state outside of the stabilizer polytope, and so its preparation is not assumed to be perfectly implementable. However, small variations in the initial error probabilities $\epsilon_i^{(0)}$ will not break the method, as illustrated in Fig. 5.

4.4 Magic as a Resource - Magic Monotones

Resource theories provide a framework for considering what is possible with a limited set of "free" operations and some set of "resources" that extend the power of these free operations. In the quantum information setting, the best-known resource theory is that of entanglement, which treats local operations and classical communication (LOCC) as free and entanglement as a resource. Many of the applications of entanglement theory concern quantum communication. Because magic states may be used to extend stabilizer computation to UQC, magic theory, in which stabilizer operations are free and magic states carry the resource, is relevant to the discussion of quantum computation [3].

An important type of tool in any resource theory is the *monotone*. A resource monotone is a function from quantum states to the reals that indicates the amount of resource possessed by the state. The only constraint on such a function is that it be non-increasing under free operations, i.e., $\mathcal{M}(\Lambda(\rho)) \leq \mathcal{M}(\rho)$ for any state ρ and any free operation Λ. However, other properties, such as additivity under the tensor product, may be desirable in order for the monotone to be a useful tool for the characterization of states. Magic monotones are defined as follows in Ref. [3]:

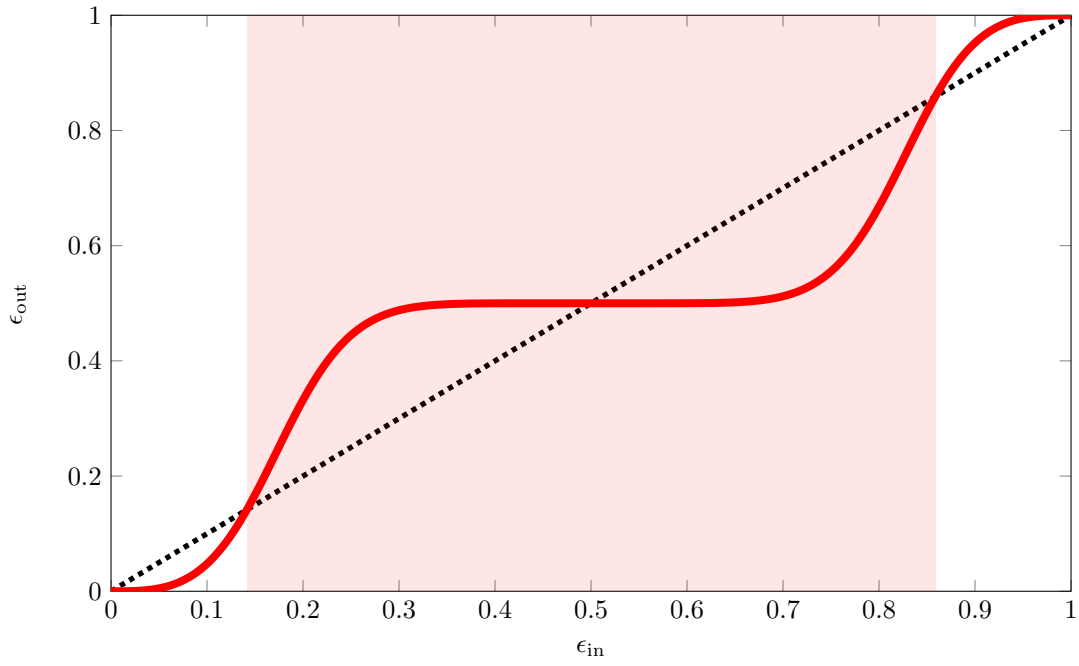
Definition 8. A map $\mathcal{M}$ from the set of all density operators to the reals is a magic monotone if and only if on average $\mathcal{M}\left(\Lambda(\rho)\right) \leq \mathcal{M}(\rho)$ for any stabilizer protocol Λ.

In particular, this means that $\mathcal{M}$ must be invariant under Clifford unitaries and tensoring-in of stabilizer states (because both of these operations have inverses that are also stabilizer operations) and non-increasing under partial trace and non-increasing on average under stabilizer measurement. The "on average" allows for magic to increase under post-selected measurement, i.e. "cheating". Note that magic may of course decrease under stabilizer protocols. For example, measuring in any stabilizer basis will decrease magic unless the pre-measurement state already had the minimum amount. While it is not required by the definition, it is convenient to have $\mathcal{M}(\rho) = 0$ for all $\rho \in \mathcal{P}_{\text{stab}}$. If this is not the case, there is a trivially related monotone $\mathcal{M}'(\rho) = \mathcal{M}(\rho) - \mathcal{M}(|0\rangle\langle 0|)$ for which it is.

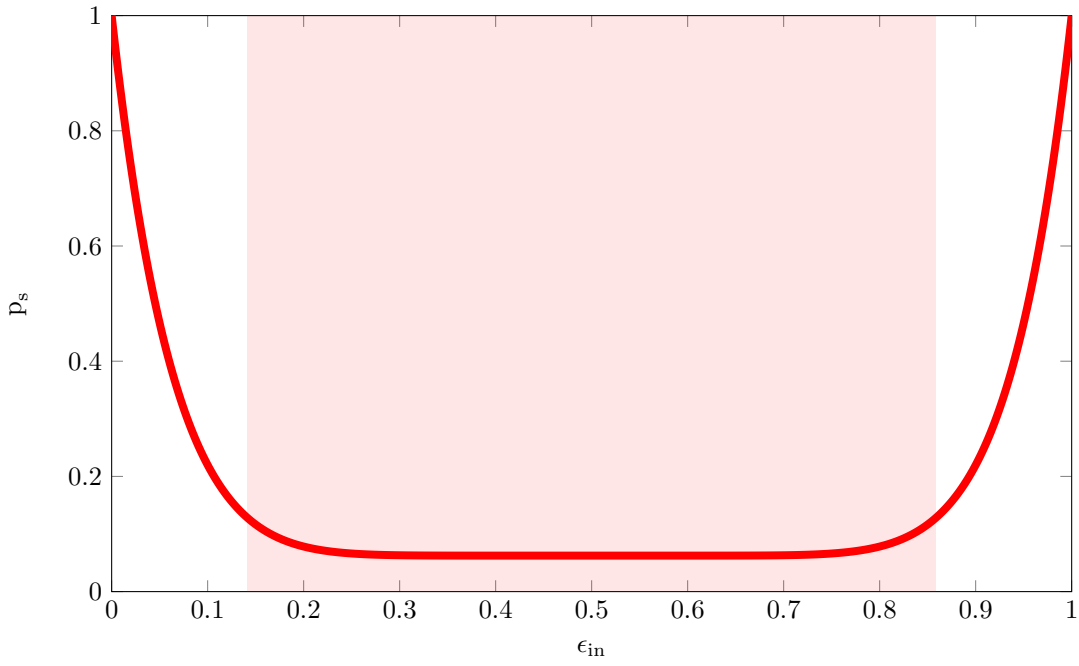
Magic monotones may be used to examine magic resource theory both qualitatively and quantitatively. They may be used to compare the efficiency of distillation protocols by comparing the ratios $\mathcal{M}(\rho_{\text{out}})/\mathcal{M}(\rho_{\text{in}})$ for each protocol. They may also, particularly when regularized, be used to determine what states are useful resource states. A regularized measure is defined [3] as follows:

$$\mathcal{M}_\infty(\rho) := \lim_{n\to\infty} \frac{1}{n}\mathcal{M}(\rho^{\otimes n}). \tag{50}$$

For additive monotones, the regularized measure is the same as the unregularized measure. For non-additive monotones, however, it is this regularized version that is the relevant quantity for the discussion of distillation.



(a)



(b)

Figure 4: In (a), output error probability ϵ_{out} is plotted against input error probability ϵ_{in}. When $\epsilon_{\text{out}} < \epsilon_{\text{in}}$, the distillation works (the error probability decreases), and when $\epsilon_{\text{out}} > \epsilon_{\text{in}}$, the distillation fails (the error probability increases). The white region on the left is the region of initial error probabilities for which the distillation yields the state $|A_0\rangle$ in the asymptotic limit. The white region on the right is the region that yields $|A_1\rangle$. The shaded region yields the fully mixed state. In (b), distillation subroutine success probability p_s is plotted against input error probability ϵ_{in}. For $\epsilon_{\text{in}} \to 0$, $p_s \to 1$, making distillation possible.

A few magic monotones have already been defined. One natural way to go about defining these functions is to look at distance measures on quantum state space. It is plausible to suppose that the minimum distance of a state from the stabilizer polytope will give a measure of magic. This is the type of monotone that will be introduced in the next section. It is also the class of monotones into which the *relative entropy of magic* [3] falls. This monotone is defined simply as the minimum relative entropy distance of a quantum state to any state in the stabilizer polytope, where the relative entropy distance between states ρ and σ is

$$S(\rho\|\sigma) := \mathrm{Tr}\left[\rho \log \rho\right] - \mathrm{Tr}\left[\rho \log \sigma\right]. \tag{51}$$

Unfortunately, no analytic expression is known for this monotone or its regularized version, and numerical evaluation quickly becomes intractible as the dimension of the Hilbert space increases.

Two more closely related magic monotones, also introduced in Ref. [3], are the *sum negativity* and the *mana*. The sum negativity $\mathrm{sn}(\rho)$ of a state ρ is defined as the sum of the absolute values of the negative elements of the Wigner function of ρ. The Wigner function is a quasiprobability representation of quantum states (see Appendices D and C). The mana is defined as

$$\mathcal{M}(\rho) := \log(2\mathrm{sn}(\rho) + 1), \tag{52}$$

which gives it the nice property of being additive:

$$\mathcal{M}(\rho \otimes \sigma) = \mathcal{M}(\rho) + \mathcal{M}(\sigma). \tag{53}$$

While the mana is simple to compute and additive, it is unfortunately undefined for qubits because the Wigner function, if positive, is a noncontextual hidden variable theory, and it is possible to violate contextuality inequalities with qubit stabilizer states [1,3,19]. For more discussion of the relationship between contextuality and quasiprobability negativity, see Appendix C.

Given the drawbacks of the existing magic monotones, a highly desirable contribution to the theory of magic would be a computable magic monotone for qubits.

4.5 A Monotone Built from a Quantum State Distance

One way to define a magic monotone is in terms of the minimum distance from a state to the convex hull of stabilizer states. There are of course many distances one can define on the space of quantum states. The Bures distance D_B provides a measure of how far apart two quantum states are in terms of their fidelity F:

$$D_B(\rho, \sigma) = \sqrt{2 - 2F(\rho, \sigma)} \qquad F(\rho, \sigma) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right] \tag{54}$$

The fidelity (Bures distance) is symmetric in ρ and σ, invariant under unitary operations, and non-decreasing (non-increasing) under trace-preserving quantum operations. Note that when $\rho = \sigma$, $F(\rho, \sigma) = 1$ and $D_B(\rho, \sigma) = 0$. When ρ and σ have orthogonal support, they may be simultaneously diagonalized, and we see that $F(\rho, \sigma) = 0$ and $D_B(\rho, \sigma) = \sqrt{2}$. For any pair (ρ, σ), we have $0 \leq F(\rho, \sigma) \leq 1$ and $0 \leq D_B(\rho, \sigma) \leq \sqrt{2}$. It will also come in handy to know that the fidelity has a nice multiplicative property:

$$F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = F(\rho_1, \sigma_1) F(\rho_2, \sigma_2) \tag{55}$$

We now define a magic monotone based on the Bures distance.

Theorem 15. *The following expression defines a magic monotone.*

$$\Upsilon(\rho) = \min_{\sigma \in \mathrm{stab}} D_B(\rho, \sigma) \tag{56}$$

where by stab_A*, we denote the convex hull of the stabilizer states of system A. When the subscript is omitted, the reference is understood to be to the same space in which* ρ *lives.*

Proof. We prove each of the requirements for magic monotonicity in turn. We will make use of two useful facts about these sets of states:

$$\sigma_1 \in \mathrm{stab}_A, \sigma_2 \in \mathrm{stab}_B \longrightarrow \sigma_1 \otimes \sigma_2 \in \mathrm{stab}_{AB} \tag{57}$$

$$\Sigma \in \mathrm{stab}_{AB} \longrightarrow \mathrm{tr}_B \Sigma \in \mathrm{stab}_A \tag{58}$$

1. Invariant under Clifford operations:

$$\Upsilon(C\rho C^\dagger) = \min_{\sigma \in \mathrm{stab}} D_B(C\rho C^\dagger, \sigma) = \min_{\sigma \in \mathrm{stab}} D_B(C\rho C^\dagger, CC^\dagger \sigma CC^\dagger) = \min_{\sigma' \in \mathrm{stab}} D_B(C\rho C^\dagger, C\sigma' C^\dagger) \tag{59}$$

$$= \min_{\sigma' \in \mathrm{stab}} D_B(\rho, \sigma') = \Upsilon(\rho) \tag{60}$$

Here we've used the fact that the Clifford unitaries map stabilizers surjectively to stabilizers.

2. Non-increasing under partial trace:

Suppose the converse. Then for some ρ, $\Upsilon(\mathrm{Tr}_B \rho) > \Upsilon(\rho)$. Substituting the definition of Υ, we have $\min_{\sigma \in \mathrm{stab}_A} D_B(\mathrm{tr}_B \rho, \sigma) > \min_{\Sigma \in \mathrm{stab}_{AB}} D_B(\rho, \Sigma)$. Then there is some $\Sigma \in \mathrm{stab}_{AB}$ such that for all $\sigma \in \mathrm{stab}_A$, $D_B(\rho, \Sigma) < D_B(\mathrm{tr}_B \rho, \sigma)$. We know that $\mathrm{tr}_B \Sigma \in \mathrm{stab}_A$, so $D_B(\mathrm{tr}_B \rho, \mathrm{tr}_B \Sigma) > D_B(\rho, \Sigma)$. But the Bures distance is non-increasing under partial trace, which is a trace-preserving quantum operation, so this is a contradiction.

3. Non-increasing on average under stabilizer measurement:

We can implement any stabilizer measurement via computational basis measurement on the final qudit if we first apply an appropriate Clifford operator, make the measurement, and then apply the inverse Clifford operator. Define the set of projectors $V_i = \mathbb{1} \otimes |i\rangle\langle i|$, un-normalized post-measurement states $\rho_i = V_i \rho V_i^\dagger$ and $\sigma_i = V_i \sigma V_i^\dagger$, and outcome probabilities $p_i = \mathrm{Tr}\left[V_i \rho\right]$ and $q_i = \mathrm{Tr}\left[V_i \sigma\right]$. The average post-measurement value of Υ is:

$$\bar{\Upsilon}(\rho_{\mathrm{post}}) = \sum_i p_i \Upsilon\left(\frac{\rho_i}{p_i}\right) = \sum_i p_i \min_{\sigma \in \mathrm{stab}} D_B\left(\frac{\rho_i}{p_i}, \sigma\right) \overset{1}{=} \sum_i p_i \min_{\sigma \in \mathrm{stab}} D_B\left(\frac{\rho_i}{p_i}, \frac{\sigma_i}{q_i}\right) \tag{61}$$

$$\leq \min_{\sigma \in \mathrm{stab}} \sum_i p_i D_B\left(\frac{\rho_i}{p_i}, \frac{\sigma_i}{q_i}\right) \overset{2}{\leq} \min_{\sigma \in \mathrm{stab}} D_B(\rho, \sigma) = \Upsilon(\rho) \tag{62}$$

Here, equality (1) follows from the fact that a post-measurement state σ_j for $j \neq i$ has orthogonal support to the post-measurement state ρ_i, so maximizes the distance. Therefore, the distance must be minimized by the post-measurement state σ_i. Inequality (2) is given in [4].

4. Invariant under tensoring-in of stabilizer states:

One direction we get for free from the partial trace behavior: $\Upsilon(\rho) = \Upsilon(\mathrm{tr}_B(\rho \otimes \tilde{\sigma})) \leq \Upsilon(\rho \otimes \tilde{\sigma})$. To show the other direction:

$$\Upsilon(\rho \otimes \tilde{\sigma}) = \min_{\Sigma \in \mathrm{stab}_{AB}} D_B(\rho \otimes \tilde{\sigma}, \Sigma) = \min_{\Sigma \in \mathrm{stab}_{AB}} \left[2 - 2F(\rho \otimes \tilde{\sigma}, \Sigma)\right]^{1/2} \tag{63}$$

$$\leq \min_{\substack{\sigma_1 \in \mathrm{stab}_A \\ \sigma_2 \in \mathrm{stab}_B}} \left[2 - 2F(\rho \otimes \tilde{\sigma}, \sigma_1 \otimes \sigma_2)\right]^{1/2} = \min_{\substack{\sigma_1 \in \mathrm{stab}_A \\ \sigma_2 \in \mathrm{stab}_B}} \left[2 - 2F(\rho, \sigma_1) F(\tilde{\sigma}, \sigma_2)\right]^{1/2} \tag{64}$$

$$\overset{1}{=} \min_{\sigma_1 \in \mathrm{stab}_A} \left[2 - 2F(\rho, \sigma_1)\right]^{1/2} = \min_{\sigma_1 \in \mathrm{stab}_A} D_B(\rho, \sigma_1) = \Upsilon(\rho) \tag{65}$$

Here, equality (1) follows from the fact that we may choose σ_2 to be a stabilizer state with orthogonal support to $\tilde{\sigma}$, saturating the fidelity $F(\tilde{\sigma}, \sigma_2)$ and thereby performing the minimization over σ_2. □

Unfortunately, this monotone does not have many especially nice properties. In particular, it is not additive or easily computable. Therfore, it will probably not be useful for practical purposes. However, the proof that it is in fact a monotone has hopefully illustrated some of the techniques that may be used in the treatment of magic theory.

5 Conclusion

In this essay, we have reviewed stabilizer computation, a subtheory of quantum computation with an efficient classical simulation algorithm. We have also presented new results about the geometric and combinatorial nature of the stabilizer polytope. In addition, we discussed magic state distillation, a method for performing fault-tolerant universal quantum computation via fault-tolerant stabilizer computation and the imperfect preparation of resource states, and introduced a new magic monotone. In the future, it would be useful to develop additive monotones for n-qubit states. It is hoped that further insight into the structure of the stabilizer polytope, as might be afforded by a complete characterization of its faces, would allow the design of a natural and simply computable monotone. Such a discovery would facilitate the examination of new magic state distillation protocols, leading to potential fault-tolerant implementations of quantum computing, and perhaps offer insight into the power of quantum computation.

A Distillation Protocol Calculations

The calculations in this appendix simply present explicitly the identities given in [2].

Lemma 16. *For an arbitrary vector $w \in \{0,1\}^{15}$, the projection of the state $|A_w\rangle$ onto the code subspace $\mathcal{H}_{ZA}(0,0)$ is*

$$\Pi|A_w\rangle = \begin{cases} 0 & w \notin \mathcal{L}_1^\perp \\ |\mathcal{L}_2|^{-\frac{1}{2}} \left|A_0^L\right\rangle & w \in \mathcal{L}_2 \\ |\mathcal{L}_2|^{-\frac{1}{2}} \left|A_1^L\right\rangle & w \in \mathcal{L}_2 \oplus [1] \end{cases} \tag{66}$$

Proof. Because the groups $A(\mathcal{L}_1)$ and $Z(\mathcal{L}_2)$ are Abelian groups with all elements squaring to identity, the projector onto the subspace $\mathcal{H}_{ZA}(0,0)$ is simply the sum of the operators that stabilize the subspace divided by the size number of such operators. Then the projection of an arbitrary state $|A\rangle_w$ onto this subspace is

$$\Pi|A_w\rangle = (|\mathcal{L}_1|\,|\mathcal{L}_2|)^{-1} \sum_{(u,v)\in\mathcal{L}_1\times\mathcal{L}_2} Z(v)A(u)|A_w\rangle \tag{67}$$

$$= (|\mathcal{L}_1|\,|\mathcal{L}_2|)^{-1} \sum_{(u,v)\in\mathcal{L}_1\times\mathcal{L}_2} (-1)^{(u,w+v)}|A_{w+v}\rangle \tag{68}$$

$$\stackrel{1}{=} (|\mathcal{L}_1|\,|\mathcal{L}_2|)^{-1} \sum_{(u,v)\in\mathcal{L}_1\times\mathcal{L}_2} (-1)^{(u,w)}|A_{w+v}\rangle \tag{69}$$

Equality (1) follows from the orthogonality of $\mathcal{L}_1$ and $\mathcal{L}_2$. Now consider the case $w \notin \mathcal{L}_1^\perp$. Then for at least one $u \in \mathcal{L}_1$, $(u,w) \equiv 1 \mod 2$. Then this is true for exactly half of the elements of $\mathcal{L}_1$. Therefore the sum over $u \in \mathcal{L}_1$ is zero, and we find that $|A_w\rangle$ is orthogonal to the subspace $\mathcal{H}_{ZA}$. In the case $w \in \mathcal{L}_1^\perp$ there are two subcases: $w \in \mathcal{L}_2 \oplus [0/1]$. Then the expression above simplifies further:

$$\stackrel{2}{=} (|\mathcal{L}_1|\,|\mathcal{L}_2|)^{-1} \sum_{(u,v)\in\mathcal{L}_1\times\mathcal{L}_2} \left|A_{v+[0/1]}\right\rangle \tag{70}$$

$$= |\mathcal{L}_2|^{-1} \sum_{v\in\mathcal{L}_2} \left|A_{v+[0/1]}\right\rangle \tag{71}$$

$$= |\mathcal{L}_2|^{-\frac{1}{2}} \left|A_{0/1}^L\right\rangle \tag{72}$$

Equality (2) follows from the coset structure of $\mathcal{L}_1^\perp$. □

The projection of the dephased input state ρ_{in} onto the code subspace $\mathcal{H}_{ZA}(0,0)$ (the post-selected post-

measurement state) is given by

$$\rho_{\rm s} = \Pi\rho_{\rm in}\Pi \tag{73}$$
$$= \sum_{u\in\{0,1\}^{15}} \epsilon^{|u|}(1-\epsilon)^{15-|u|}\Pi|A_u\rangle\langle A_u|\Pi \tag{74}$$
$$= \sum_{u\in\mathcal{L}_2} \epsilon^{|u|}(1-\epsilon)^{15-|u|}\Pi|A_u\rangle\langle A_u|\Pi + \sum_{u\in\mathcal{L}_2+[1]} \epsilon^{|u|}(1-\epsilon)^{15-|u|}\Pi|A_u\rangle\langle A_u|\Pi \tag{75}$$
$$= \sum_{u\in\mathcal{L}_2} \epsilon^{|u|}(1-\epsilon)^{15-|u|}\left|\mathcal{L}_2\right|^{-1}\left|A_0^L\right\rangle\left\langle A_0^L\right| + \sum_{u\in\mathcal{L}_2+[1]} \epsilon^{|u|}(1-\epsilon)^{15-|u|}\left|\mathcal{L}_2\right|^{-1}\left|A_1^L\right\rangle\left\langle A_1^L\right| \tag{76}$$
$$= \left|\mathcal{L}_2\right|^{-1}\sum_{u\in\mathcal{L}_2} \epsilon^{|u|}(1-\epsilon)^{15-|u|}\left|A_0^L\right\rangle\left\langle A_0^L\right| + \left|\mathcal{L}_2\right|^{-1}\sum_{u\in\mathcal{L}_2} \epsilon^{15-|u|}(1-\epsilon)^{|u|}\left|A_1^L\right\rangle\left\langle A_1^L\right| \tag{77}$$

It is convenient to expand the basis states of the code subspace $\mathcal{H}_{ZA}(0,0)$ in the computational basis:

$$\left|A_0^L\right\rangle = \left|\mathcal{L}_2\right|^{-\frac{1}{2}}\sum_{v\in\mathcal{L}_2}|A_v\rangle \tag{78}$$
$$= \left|\mathcal{L}_2\right|^{-\frac{1}{2}}\sum_{v\in\mathcal{L}_2} Z(v)\left|A_{[0]}\right\rangle \tag{79}$$
$$= \left|\mathcal{L}_2\right|^{-\frac{1}{2}}\sum_{v\in\mathcal{L}_2} Z(v)2^{-\frac{15}{2}}\left(|0\rangle + e^{\frac{i\pi}{4}}|1\rangle\right)^{\otimes 15} \tag{80}$$
$$= 2^{-\frac{15}{2}}\left|\mathcal{L}_2\right|^{-\frac{1}{2}}\sum_{v\in\mathcal{L}_2}\sum_{u\in\{0,1\}^{15}} e^{\frac{i\pi}{4}|u|}Z(v)|u\rangle \tag{81}$$
$$= 2^{-\frac{15}{2}}\left|\mathcal{L}_2\right|^{-\frac{1}{2}}\sum_{v\in\mathcal{L}_2}\sum_{u\in\{0,1\}^{15}} e^{\frac{i\pi}{4}|u|}(-1)^{(u,v)}|u\rangle \tag{82}$$
$$\stackrel{1}{=} 2^{-\frac{15}{2}}\left|\mathcal{L}_2\right|^{-\frac{1}{2}}\sum_{v\in\mathcal{L}_2}\sum_{u\in\mathcal{L}_2^\perp} e^{\frac{i\pi}{4}|u|}|u\rangle \tag{83}$$
$$= 2^{-\frac{15}{2}}\left|\mathcal{L}_2\right|^{\frac{1}{2}}\sum_{u\in(\mathcal{L}_2)^\perp} e^{\frac{i\pi}{4}|u|}|u\rangle \tag{84}$$
$$= 2^{-\frac{15}{2}}\left|\mathcal{L}_2\right|^{\frac{1}{2}}\left(\sum_{u\in\mathcal{L}_1} e^{\frac{i\pi}{4}|u|}|u\rangle + \sum_{u\in\mathcal{L}_1+[1]} e^{\frac{i\pi}{4}|u|}|u\rangle\right) \tag{85}$$
$$\stackrel{2}{=} 2^{-\frac{5}{2}}\sum_{u\in\mathcal{L}_1}\left(|u\rangle + e^{-\frac{i\pi}{4}}|u+[1]\rangle\right) \tag{86}$$

Equality (1) follows from the fact, explained earlier, that for a fixed $u \notin \mathcal{L}_2^\perp$, exactly half of the elements $v \in \mathcal{L}_2$ have even zero product with u, so the sum over v vanishes. Equality (2) follows from the fact that for all $u \in \mathcal{L}_1$, $|u| \equiv 0 \mod 8$. Now we can compute the expectation values of the logical Pauli operators

for the basis states:

$$\left\langle A_0^L\right|X_L\left|A_0^L\right\rangle = 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left(\langle u| + e^{\frac{i\pi}{4}}\langle u+[1]|\right)X^{\otimes 15}\left(|v\rangle + e^{-\frac{i\pi}{4}}|v+[1]\rangle\right) \tag{87}$$

$$= 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[\langle u|\overline{v}\rangle + e^{-\frac{i\pi}{4}}\left\langle u|\overline{v+[1]}\right\rangle + e^{\frac{i\pi}{4}}\langle u+[1]\,|\overline{v}\rangle + \left\langle u+1|\overline{v+[1]}\right\rangle\right] \tag{88}$$

$$\overset{1}{=} 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[e^{-\frac{i\pi}{4}}\left\langle u|\overline{v+[1]}\right\rangle + e^{\frac{i\pi}{4}}\langle u+[1]\,|\overline{v}\rangle\right] \tag{89}$$

$$= 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[\left(e^{-\frac{i\pi}{4}} + e^{\frac{i\pi}{4}}\right)\langle u|v\rangle\right] \tag{90}$$

$$= \frac{2^{-4}}{\sqrt{2}}\,|\mathcal{L}_1| \tag{91}$$

$$= \frac{1}{\sqrt{2}} \tag{92}$$

$$\left\langle A_0^L\right|Y_L\left|A_0^L\right\rangle = 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left(\langle u| + e^{\frac{i\pi}{4}}\langle u+[1]|\right)Y^{\otimes 15}\left(|v\rangle + e^{-\frac{i\pi}{4}}|v+[1]\rangle\right) \tag{93}$$

$$= 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[e^{-\frac{i\pi}{4}}\langle u|Y^{\otimes 15}|v+[1]\rangle + e^{\frac{i\pi}{4}}\langle u+[1]|Y^{\otimes 15}|v\rangle\right] \tag{94}$$

$$= 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[e^{-\frac{i\pi}{4}}\langle u|(-i)(-1)^{|v+[1]|}\left|\overline{v+[1]}\right\rangle + e^{\frac{i\pi}{4}}\langle u+[1]|(-i)(-1)^{|v|}|\overline{v}\rangle\right] \tag{95}$$

$$= -i2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[e^{-\frac{i\pi}{4}}(-1)^{|v+[1]|}\langle u|v\rangle + e^{\frac{i\pi}{4}}(-1)^{|v|}\langle u+[1]\,|\overline{v}\rangle\right] \tag{96}$$

$$= -i2^{-5}\sum_{u\in\mathcal{L}_1}\left[e^{-\frac{i\pi}{4}}(-1)^{|u+[1]|} + e^{\frac{i\pi}{4}}(-1)^{|u|}\right] \tag{97}$$

$$= -i2^{-5}\sum_{u\in\mathcal{L}_1}\left[-e^{-\frac{i\pi}{4}} + e^{\frac{i\pi}{4}}\right] \tag{98}$$

$$= \frac{1}{\sqrt{2}} \tag{99}$$

$$\left\langle A_0^L\right|Z_L\left|A_0^L\right\rangle = 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left(\langle u| + e^{\frac{i\pi}{4}}\langle u+[1]|\right)\left(-Z^{\otimes 15}\right)\left(|v\rangle + e^{-\frac{i\pi}{4}}|v+[1]\rangle\right) \tag{100}$$

$$= 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[-(-1)^{|v|}\langle u|v\rangle - (-1)^{|v+[1]|}\langle u+[1]\,|v+[1]\rangle\right] \tag{101}$$

$$= 2^{-5}\sum_{u,v\in\mathcal{L}_1}\left[\langle u+[1]\,|v+[1]\rangle - \langle u|v\rangle\right] \tag{102}$$

$$= 0 \tag{103}$$

Frequent use is made of the fact that $u+[1] = \overline{u}$. Equality (1) follows from the fact that $[1] \notin \mathcal{L}_1$.

To demonstrate that $\left|A_0^L\right\rangle$ is indeed the logical state corresponding to the single-qubit state $|A_0\rangle$, we simply need to show that the expectations of the single-qubit Pauli operators have the same expectations with $|A_0\rangle$ as the logical Paulis do with $\left|A_0^L\right\rangle$:

$$\langle A_0|X|A_0\rangle = \frac{1}{2}\left(\langle 0| + e^{-\frac{i\pi}{4}}\langle 1|\right) X \left(|0\rangle + e^{\frac{i\pi}{4}}|1\rangle\right) \tag{104}$$
$$= \frac{1}{2}\left(\langle 0| + e^{-\frac{i\pi}{4}}\langle 1|\right)\left(|1\rangle + e^{\frac{i\pi}{4}}|0\rangle\right) \tag{105}$$
$$= \frac{1}{2}\left(e^{\frac{i\pi}{4}} + e^{-\frac{i\pi}{4}}\right) \tag{106}$$
$$= \frac{1}{\sqrt{2}} \tag{107}$$
$$\langle A_0|Y|A_0\rangle = \frac{1}{2}\left(\langle 0| + e^{-\frac{i\pi}{4}}\langle 1|\right) Y \left(|0\rangle + e^{\frac{i\pi}{4}}|1\rangle\right) \tag{108}$$
$$= \frac{1}{2}\left(\langle 0| + e^{-\frac{i\pi}{4}}\langle 1|\right)\left(i|1\rangle - ie^{\frac{i\pi}{4}}|0\rangle\right) \tag{109}$$
$$= -\frac{i}{2}\left(e^{i\frac{\pi}{4}} - e^{-i\frac{\pi}{4}}\right) \tag{110}$$
$$= \frac{1}{\sqrt{2}} \tag{111}$$
$$\langle A_0|Z|A_0\rangle = \frac{1}{2}\left(\langle 0| + e^{-\frac{i\pi}{4}}\langle 1|\right) Z \left(|0\rangle + e^{\frac{i\pi}{4}}|1\rangle\right) \tag{112}$$
$$= \frac{1}{2}\left(\langle 0| + e^{-\frac{i\pi}{4}}\langle 1|\right)\left(|0\rangle - e^{\frac{i\pi}{4}}|1\rangle\right) \tag{113}$$
$$= 0 \tag{114}$$

Because in a two-dimensional subspace there is only one state orthogonal to any given state and because unitary channels preserve orthogonality, this also tells us that $|A_1\rangle$ is the logical state corresponding to the single-qubit state $|A_1\rangle$.

B Generalizing Stabilizer Protocols to Qudits

Although the discussion in the body of the essay has focused solely on systems of qubits, it is also possible to consider systems of qudits, quantum systems with dimension d^n for d an odd prime. Stabilizer computation, defined analogously to the qubit case, is also efficiently simulable in the qudit case. The following discussion follows that of [3]. We wish to generalize the qubit Pauli operators to operators acting on Hilbert spaces of odd prime dimension. To this end, we define generators

$$X|j\rangle = |j+1\rangle \tag{115}$$
$$Z|j\rangle = e^{\frac{2\pi i}{d}j}|j\rangle \tag{116}$$

Throughout this discussion, arithmetic is understood to be modulo d. These definitions hold as well for $d = 2$. For odd prime dimension, we now define the single-qudit Heisenberg-Weyl operators

$$T_{(a,b)} = e^{-\frac{\pi i}{d}ab} Z^a X^b \tag{117}$$

where $a, b = 1, \ldots, d-1$. Note that these operators are not in general Hermitian.

Each qudit operator may be described by an ordered pair (a_i, b_i), so we may describe the operator on the n-qudit space by the vector $u = (\vec{u}_a, \vec{u}_b) = (a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n) \in \mathbb{Z}_d^{2n}$ and denote the operators T_u. These operators obey some nice algebraic relations (demonstrated below):

$$T_u T_v = e^{\frac{\pi i}{d}\langle u,v\rangle} T_{u+v} \tag{118}$$

$$[T_u, T_v] = 2i \sin\left[\frac{\pi}{d}\langle u, v\rangle\right] T_{u+v} \tag{119}$$

$$T_u^\dagger = T_{-u} \tag{120}$$

$$T_u T_v T_u^\dagger = e^{\frac{2\pi i}{d}\langle u,v\rangle} T_v \tag{121}$$

where for $u = (\vec{u}_a, \vec{u}_b)$ and $v = (\vec{v}_a, \vec{v}_b)$, $\langle u, v\rangle$ is the symplectic inner product

$$\langle u, v\rangle = (\vec{u}_a, \vec{u}_b) \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \vec{v}_a^T \\ \vec{v}_b^T \end{pmatrix}. \tag{122}$$

Note that $\langle u, u\rangle = 0$ for any vector u.

The Heisenberg-Weyl operators are mutually orthogonal with respect to the Hilbert-Schmidt operator:

$$\mathrm{Tr}\left[T_u^\dagger T_v\right] = d^n \delta_{u,v} \tag{123}$$

The operators are specified by $2n$-dimensional vectors with integer entries between 0 and $d-1$, so there are d^{2n} such operators. Operators on the Hilbert space of an n-qudit system are $d^n \times d^n$ matrices, so are spanned by d^n basis operators. Therefore, the Heisenberg-Weyl operators are an orthogonal basis for operators on $\mathcal{H}_{n-\mathrm{qudit}}$. Then, we can write any operator in the form

$$\rho = \frac{1}{d^n} \sum_u r_u T_u \qquad r_u = \mathrm{Tr}\left[T_u^\dagger \rho\right] \tag{124}$$

If ρ is a density operator, we must have $r_0 = 1$ to satisfy the trace one condition and $r_{-u} = r_u^*$ to satisfy Hermiticity. Now we have a nice way to represent arbitrary n-qudit states in vector form. Of course, this representation is still exponentially large, as expected. However, as in the qubit case, we may represent efficiently those states that are stabilized by Abelian subgroups of the Heisenberg-Weyl group with polynomially large matrices with elements from the field F_d, and stabilizer operations on these states may be simulated efficiently with protocols analogous to those given for the qubit case in the body of this essay. Some technical details about the group are given below.

B.1 Algebraic Properties of the Heisenberg-Weyl Operators

We can express the operator $T_{(a,b)}$ explicitly as follows:

$$T_{(a,b)} = e^{-\frac{\pi i}{d}ab} Z^a X^b = e^{-\frac{\pi i}{d}ab} \left(\sum_{j=0}^{d-1} e^{\frac{2\pi i}{d}j} |j\rangle\langle j|\right)^a \left(\sum_{k=0}^{d-1} |k\rangle\langle k -_d 1|\right)^b \tag{125}$$

$$= e^{-\frac{\pi i}{d}ab} \left(\sum_{j=0}^{d-1} e^{\frac{2\pi i}{d}ja} |j\rangle\langle j|\right) \left(\sum_{k=0}^{d-1} |k\rangle\langle k -_d b|\right) = e^{-\frac{\pi i}{d}ab} \sum_{j=0}^{d-1} e^{\frac{2\pi i}{d}ja} |j\rangle\langle j -_d b| \tag{126}$$

$$= e^{-\frac{\pi i}{d}ab} \sum_{j=0}^{d-1} e^{\frac{2\pi i}{d}(j+_d b)a} |j +_d b\rangle\langle j| = e^{\frac{\pi i}{d}ab} \sum_{j=0}^{d-1} e^{\frac{2\pi i}{d}ja} |j +_d b\rangle\langle j| \tag{127}$$

It will come in handy to know how to pull pure X and Z operators past each other:

$$X^a Z^b = \left(\sum_{j=0}^{d-1} |j +_d a\rangle\langle j|\right) \left(\sum_{k=0}^{d-1} e^{\frac{2\pi i}{d}kb} |k\rangle\langle k|\right) = \sum_{j=0}^{d-1} e^{\frac{2\pi i}{d}jb} |j +_d a\rangle\langle j| = e^{-\frac{\pi i}{d}ab} T_{(b,a)} = e^{-\frac{2\pi i}{d}ab} Z^b X^a \tag{128}$$

Single-qudit operators compose in the following fashion:

$$T_{(a,b)} T_{(c,d)} = e^{-\frac{\pi i}{d}ab} Z^a X^b e^{-\frac{\pi i}{d}cd} Z^c X^d = e^{-\frac{\pi i}{d}(ab+cd)} Z^a X^b Z^c X^d \tag{129}$$

$$= e^{-\frac{\pi i}{d}(ab+cd)} e^{-\frac{2\pi i}{d}bc} Z^a Z^c X^b X^d = e^{-\frac{\pi i}{d}(ab+cd)} e^{-\frac{2\pi i}{d}bc} Z^{a+c} X^{b+d} \tag{130}$$

$$= e^{-\frac{\pi i}{d}(ab+cd)} e^{\frac{\pi i}{d}(a+c)(b+d)} e^{-\frac{2\pi i}{d}bc} T_{(a+c,b+d)} \tag{131}$$

$$= e^{\frac{\pi i}{d}(ad+bc)} e^{-\frac{2\pi i}{d}bc} T_{(a+c,b+d)} = e^{\frac{\pi i}{d}(ad-bc)} T_{(a+c,b+d)} \tag{132}$$

To extend to the n-qudit case, define $u = (a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n)$ and $v = (c_1, c_2, \ldots, c_n, d_1, d_2, \ldots, d_n)$ Then the product of the n-qudit Heisenberg-Weyl operators specified by u and v is:

$$\begin{aligned} T_u T_v &= T_{(a_1,\ldots,a_n,b_1,\ldots,b_n)} T_{(c_1,\ldots,c_n,d_1,\ldots,d_n)} && (133)\\ &= \left(T_{(a_1,b_1)} \otimes \ldots \otimes T_{(a_n,b_n)}\right)\left(T_{(c_1,d_1)} \otimes \ldots \otimes T_{(c_n,d_n)}\right) && (134)\\ &= T_{(a_1,b_1)} T_{(c_1,d_1)} \otimes \ldots \otimes T_{(a_n,b_n)} T_{(c_n,d_n)} && (135)\\ &= e^{\frac{\pi i}{d}(a_1 d_1 - b_1 c_1)} T_{(a_1+c_1,b_1+d_1)} \otimes \ldots \otimes e^{\frac{\pi i}{d}(a_n d_n - b_n c_n)} T_{(a_n+c_n,b_n+d_n)} && (136)\\ &= e^{\sum_i \frac{\pi i}{d}(a_i d_i - b_i c_i)} T_{(a_1+c_1,\ldots,a_n+c_n,b_1+d_1,\ldots b_n+d_n)} && (137)\\ &= e^{\frac{\pi i}{d}\langle u,v\rangle} T_{u+v} && (138) \end{aligned}$$

Then the commutation relations are given by:

$$\begin{aligned} [T_u, T_v] = T_u T_v - T_v T_u &= e^{\frac{\pi i}{d}\langle u,v\rangle} T_{u+v} - e^{\frac{\pi i}{d}\langle v,u\rangle} T_{v+u} && (139)\\ &= \left(e^{\frac{\pi i}{d}\langle u,v\rangle} - e^{-\frac{\pi i}{d}\langle u,v\rangle}\right) T_{u+v} = 2i \sin\left[\frac{\pi}{d}\langle u,v\rangle\right] T_{u+v} && (140) \end{aligned}$$

The adjoint of a single-qudit Heisenberg-Weyl operator is given by:

$$\begin{aligned} T^\dagger_{(a,b)} &= e^{\frac{\pi i}{d}ab} X^{b\dagger} Z^{a\dagger} = e^{\frac{\pi i}{d}ab} X^{-b} Z^{-a} = e^{\frac{\pi i}{d}ab} e^{-\frac{2\pi i}{d}(-a)(-b)} Z^{-a} X^{-b} && (141)\\ &= e^{-\frac{\pi i}{d}(-a)(-b)} Z^{-a} X^{-b} = T_{(-a,-b)} && (142) \end{aligned}$$

The adjoint of a tensor product is simply the tensor product of adjoints, so this gives in the n-qudit case

$$T^\dagger_u = T_{-u} \tag{143}$$

Then the operators obey the conjugation relation:

$$\begin{aligned} T_u T_v T_u^\dagger = T_u T_v T_{-u} &= e^{\frac{\pi i}{d}\langle u,v\rangle} T_{u+v} T_{-u} = e^{\frac{\pi i}{d}\langle u,v\rangle} e^{\frac{\pi i}{d}\langle u+v,-u\rangle} T_v && (144)\\ &= e^{\frac{\pi i}{d}\langle u,v\rangle} e^{\frac{\pi i}{d}(\langle u,u\rangle + \langle u,v\rangle)} T_v = e^{\frac{2\pi i}{d}\langle u,v\rangle} T_v && (145) \end{aligned}$$

B.2 Representing Operators with Heisenberg-Weyl Operators

The Heisenberg-Weyl Operators are mutually orthogonal with respect to the Hilbert-Schmidt inner product:

$$\begin{aligned} \mathrm{Tr}\left[T_u^\dagger T_v\right] = \mathrm{Tr}\left[T_{-u} T_v\right] = \mathrm{Tr}\left[e^{\frac{\pi i}{d}\langle -u,v\rangle} T_{-u+v}\right] &= e^{\frac{\pi i}{d}\langle v,u\rangle} \mathrm{Tr}\left[T_{v-u}\right] && (146)\\ &= e^{\frac{\pi i}{d}\langle v,u\rangle} \mathrm{Tr}\left[T_{(c_1,\ldots,c_n d_1,\ldots,d_n)-(a_1,\ldots,a_n b_1,\ldots,b_n)}\right] && (147)\\ &= e^{\frac{\pi i}{d}\langle v,u\rangle} \mathrm{Tr}\left[T_{(c_1-a_1,d_1-b_1)} \otimes \ldots \otimes T_{(c_n-a_n,d_n-b_n)}\right] && (148)\\ &= e^{\frac{\pi i}{d}\langle v,u\rangle} \prod_{j=1}^{n} \mathrm{Tr}\left[T_{(c_j-a_j,d_j-b_j)}\right] = e^{\frac{\pi i}{d}\langle v,u\rangle} \prod_{j=1}^{n} d\delta_{c_j,a_j}\delta_{d_j,b_j} && (149)\\ &= e^{\frac{\pi i}{d}\langle v,u\rangle} d^n \delta_{u,v} = d^n \delta_{u,v} && (150) \end{aligned}$$

Then we can write an operator in the form

$$\rho = \frac{1}{d^n} \sum_u r_u T_u \tag{151}$$

This gives us a simple way to find the decomposition of an operator on $\mathcal{H}_{d^n}$ in terms of the Heisenberg-Weyl operators:

$$\mathrm{Tr}\left[T_u^\dagger \rho\right] = \mathrm{Tr}\left[T_u^\dagger \frac{1}{d^n} \sum_v r_v T_v\right] = \frac{1}{d^n} \sum_v r_v \mathrm{Tr}\left[T_u^\dagger T_v\right] = \frac{1}{d^n} \sum_v r_v d^n \delta_{u,v} = r_u \tag{152}$$

To determine the constraints on the values of r_u for a density matrix, we begin by calculating the square of the operator.

$$\rho^2 = \frac{1}{d^{2n}} \sum_{u,v} r_u r_v T_u T_v = \frac{1}{d^{2n}} \sum_{u,v} r_u r_v e^{\frac{\pi i}{d}\langle u,v\rangle} T_{u+v} = \frac{1}{d^{2n}} \sum_{s,u} r_u r_{s-u} e^{\frac{\pi i}{d}\langle u,s-u\rangle} T_s \tag{153}$$

$$= \frac{1}{d^{2n}} \sum_{s,u} r_u r_{s-u} e^{\frac{\pi i}{d}\langle u,s\rangle} T_s \tag{154}$$

Then the trace is

$$\mathrm{Tr}\left[\rho^2\right] = \frac{1}{d^{2n}} \sum_u r_u r_{-u} d^n = \frac{1}{d^n} \sum_u r_u r_u^* \tag{155}$$

A valid density operator has $\mathrm{Tr}\left[\rho^2\right] \leq \mathrm{Tr}\left[\rho\right] = 1$, with equality for pure states. Then we must have

$$\frac{1}{d^n}\left(1 + \vec{r}^\dagger \vec{r}\right) \leq 1 \tag{156}$$

where we've used the fact that $r_0 = 1$ and grouped the remaining $d^{2n} - 1$ coefficients into the vector $\vec{r}$. In the single-qubit case, $\vec{r}$ is simply the Bloch vector, and any $\vec{r}$ satisfying the above condition is valid. In higher dimensions, there are further constraints on $\vec{r}$ imposed by positivity.

C Noncontextual Ontological Model = Positive Quasiprobability Representation

Spekkens has noted in [12] that negativity of all quasiprobability representations of a quantum (sub)theory reflects the same departure from classicality as noncontextuality violation. In order to do so, he generalizes the notion of contextuality to include preparation and reject the insistence on deterministic sharp (projective) measurements. In [10], it is demonstrated that these are well-justified alterations of the concept of contextuality. A particular example is that of Bell inequality violation with EPR pairs. These states admit positive Wigner representations, but the necessary measurements have Wigner representations outside of $[0, 1]$, so cannot be interpreted as probabilities. The following definitions and argument are due to Spekkens.

Definition 9. A quasiprobability representation of quantum theory consists of the following components:

- A measurable space Λ with elements λ
- A function $\mu_\rho : \Lambda \to \mathbb{R}$ associated to each density operator ρ such that $\int d\lambda \mu_\rho(\lambda) = 1$
- A set of functions $\xi_{E_k} : \Lambda \to \mathbb{R}$ associated to each POVM $\{E_k\}$ such that $\sum_k \xi_{E_k}(\lambda) = 1$
- The constraint $\mathrm{Tr}\left[\rho E_k\right] = \int d\lambda \mu_\rho(\lambda) \xi_{E_k}(\lambda)$

A nonnegative quasiprobability representation is one in which $\mu_\rho(\lambda) \geq 0$ and $\xi_E(\lambda) \geq 0$ for all $\lambda \in \Lambda$, all density operators ρ, and all positive operators E less than identity (in the sense of the partial order $A \leq B$ if and only if $B - A$ is positive). Such operators are potential POVM elements.

An operational theory has as its primitive elements preparations, transformations, and measurements and aims to determine the probabilities $p(k|P, T, M)$ of outcome k given preparation P, transformation T, and measurement M. Equivalence classes of preparations are defined so that

$$P \sim P' \longleftrightarrow p(k|P, T, M) = p(k|P', T, M) \forall T, M \tag{157}$$

An ontological model of an operational theory is an explanation of the predictions of the theory in terms of physical systems. For convenience, we will consider only the preparation and measurement components of the theory, but transformations may be considered similarly as examined in detail in [10]. It is also possible to cast transformations in the language of quasiprobability representations [8, 19].

Definition 10. An ontological model of quantum theory consists of the following components:

- A measurable space λ with elements λ
- A function $\mu_P : \Lambda \to \mathbb{R}$ associated to each preparation P such that $\int d\lambda \mu_P(\lambda) = 1$ and $\mu_P(\lambda) \geq 0$
- A function $\xi_{M,k} : \Lambda \to \mathbb{R}$ associated to each measurement M such that $\sum_k \xi_{M,k}(\lambda) = 1$ and $\xi_{M,k}(\lambda) \geq 0$
- The constraint $\mathrm{Tr}\left[\rho E_k\right] = \int d\lambda \mu_P(\lambda) \xi_{M,k}(\lambda)$ for all P in the equivalence class $\mathcal{P}_\rho$ and all M in the equivalence class $\mathcal{M}_{\{E_k\}}$.

A noncontextual ontological model is one for which $\mu_P(\lambda) = \mu_\rho(\lambda)$ for all $P \in \mathcal{P}_\rho$ and $\xi_{M,k}(\lambda) = \xi_{E_k}(\lambda)$ for all $M \in \mathcal{M}_{\{E_k\}}$. From these definitions, it is clear that a nonnegative quasiprobability representation of quantum theory *is* a noncontextual ontological model. Therefore, we may conclude that a theory admits a noncontextual ontological model if and only if it admits a nonnegative quasiprobability representation.

The relationship between contextuality and universal quantum computation, glimsped through the lens of the mana, suggests that contextuality may be related to the posited quantum speedup [1, 7–9, 19]. This is a fascinating area for future study.

D The Discrete Wigner Function

We may define a set of single-qudit Hermitian operators, the phase space point operators, as follows:

$$A_0 = \frac{1}{d} \sum_{a,b=0}^{d-1} T_{(a,b)} \tag{158}$$

$$A_{(a,b)} = T_{(a,b)} A_0 T_{(a,b)}^\dagger \tag{159}$$

For n qudits, a and b become vectors in $\mathbb{Z}_d^n$, and we define

$$T_{(a\oplus c, b\oplus d)} = T_{(a,b)} \otimes T_{(c,d)} \tag{160}$$

From this it follows that

$$A_{(a\oplus c, b\oplus d)} = A_{(a,b)} \otimes T_{(c,d)} \tag{161}$$

We may now define the discrete Wigner function:

$$W_\rho(a,b) = \frac{1}{d^n} \mathrm{Tr}\left[A_{(a,b)}\rho\right]; \;\; a,b \in \mathbb{Z}_d^n \tag{162}$$

Because the phase space point operators $A_{(a,b)}$ are Hermitian, the Wigner function is real-valued. If $W_\rho(a,b) \geq 0$ for all a, b, the state ρ is said to have positive representation. Otherwise, it is said to have negative representation.

Lemma 17. *The discrete Wigner function is a quasiprobability representation for quantum states.*

Proof. The first requirement of a quasiprobability representation of a quantum theory is that it be defined on a measurable space Λ. In this case, $\Lambda = \{(a,b) | a, b \in \mathbb{Z}_d^n\}$. We require a function $\mu_\rho : \Lambda \to \mathbb{R}$ to be associated to each density matrix ρ such that $\int d\lambda \mu_\rho(\lambda) = 1$. This is provided by W_ρ. We have already established that this function is real-valued. Because Λ is discreet, the integral is replaced by a sum, and we have:

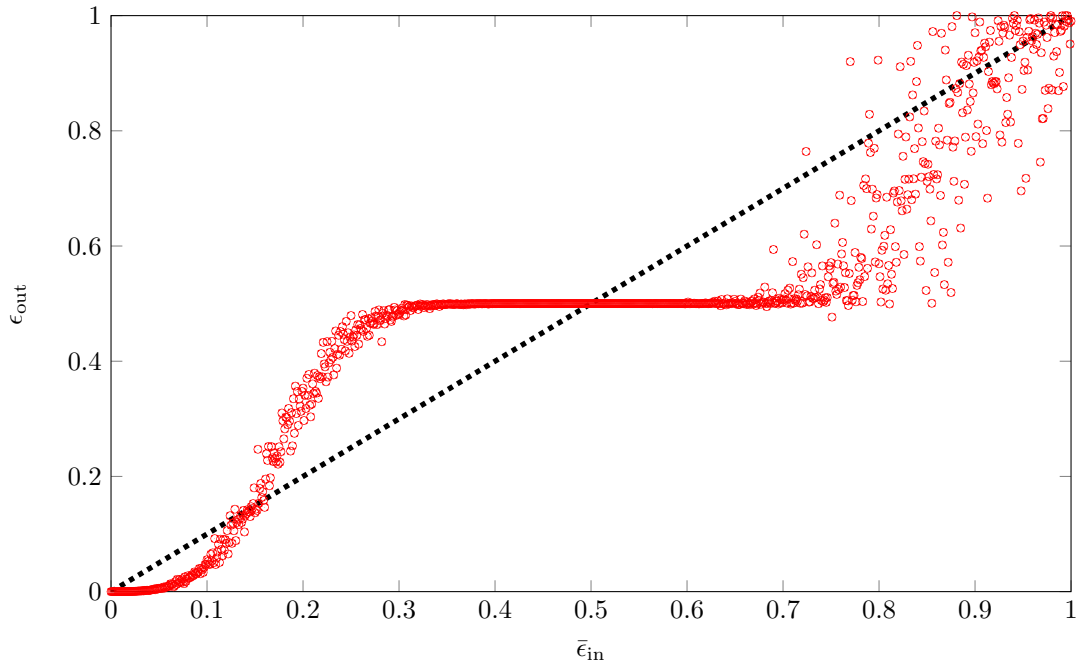
$$\sum_u W_\rho(u) = \sum_u \frac{1}{d^n} \mathrm{Tr}(A_u \rho) = \frac{1}{d^n} \mathrm{Tr}\left(\sum_u T_u A_0 T_u^\dagger \rho\right) = \frac{1}{d^{2n}} \mathrm{Tr}\left[\left(\sum_{u,v} T_u T_v T_u^\dagger\right)\rho\right] \tag{163}$$

$$= \frac{1}{d^{2n}} \mathrm{Tr}\left[\sum_v \left(\sum_u e^{\frac{2\pi i}{d}(u\times v)} T_v\right)\rho\right] = \mathrm{Tr}\left[T_0 \rho\right] = \mathrm{Tr}(\rho) = 1 \tag{164}$$

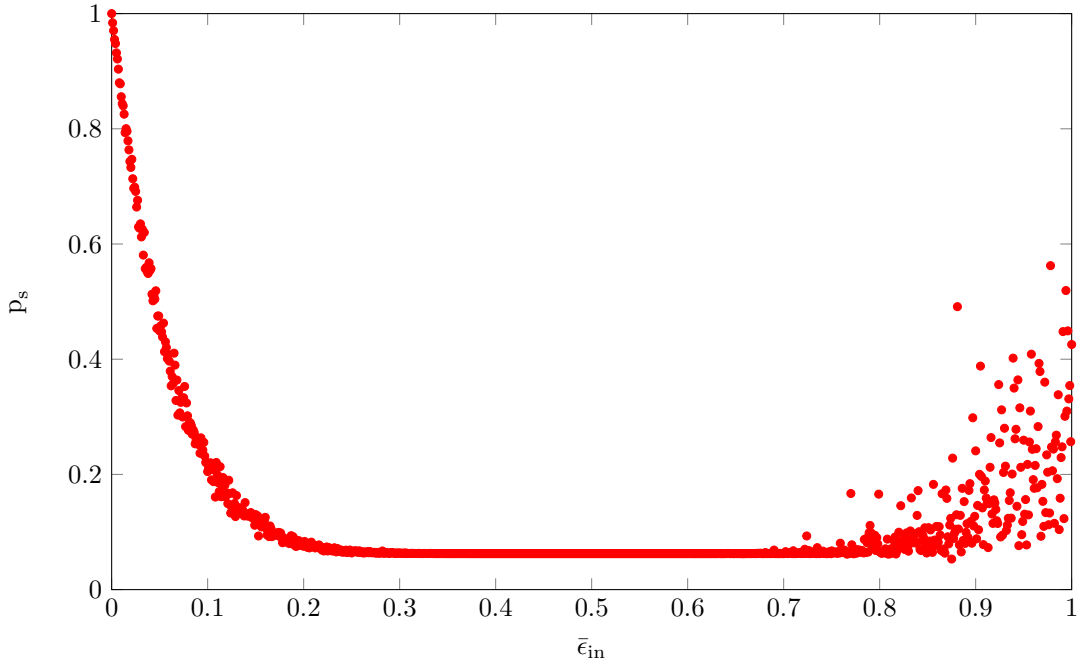
where we've used:

$$\sum_u e^{\frac{2\pi i}{d}(u\times v)} = \begin{cases} d^{2n} & v = 0 \\ 0 & v \neq 0 \end{cases}$$

□



(a)



(b)

Figure 5: In (a), output error probability ϵ_{out} is plotted against average input error probability $\bar{\epsilon}_{\text{in}}$. The error distribution used is a gaussian with mean $\bar{\epsilon}_{\text{in}}$ and standard deviation $\bar{\epsilon}_{\text{in}}/5$, truncated at 0 and 1. In (b), distillation subroutine success probability p_s is plotted against average input error probability $\bar{\epsilon}_{\text{in}}$. Of course, there is no guarantee that the form of the error distribution is invariant under the distillation protocol, but these results at least suggest that distillation is still possible even with some limited inhomogeneity in the error probabilities of the resource states $\rho^{(0)}$.

References

[1] Gross, D. (2006). Hudsons theorem for finite-dimensional quantum systems. Journal of Mathematical Physics, 47(12).

[2] Bravyi, S., Kitaev, A. (2005). Universal quantum computation with ideal Clifford gates and noisy ancillas. Physical Review A - Atomic, Molecular, and Optical Physics, 71(2).

[3] Veitch, V. Mousavian, S. A. H., Gottesman, D., Emerson, J. (2013). The Resource Theory of Stabilizer Computation. arXiv:1307.7171

[4] Vedral, V., Plenio, M. B. (1997). Entanglement measures and purification procedures. arXiv:quant-ph/9707035

[5] Gottesman, D. (1998). The Heisenberg Representation of Quantum Computers. arXiv:quant-ph/9807006

[6] Michael A. Nielsen and Isaac L. Chuang. 2011. Quantum Computation and Quantum Information: 10th Anniversary Edition (10th ed.). Cambridge University Press, New York, NY, USA.

[7] Howard, M., Wallman, J. J., Veitch, V., Emerson, J. (2014). Contextuality supplies the magic for quantum computation. arXiv preprint arXiv: , 5. Retrieved from http://arxiv.org/abs/1401.4174

[8] Cormick, C., Paz, J. P. (2006). Interference in discrete Wigner functions. Physical Review A - Atomic, Molecular, and Optical Physics, 74(6).

[9] Cabello, A., Severini, S., Winter, A. (2010). (Non-)Contextuality of Physical Theories as an Axiom. arXiv:1010.2163

[10] Spekkens, R. W. (2005). Contextuality for preparations, transformations, and unsharp measurements. Physical Review A - Atomic, Molecular, and Optical Physics, 71(5).

[11] Aaronson, S., Gottesman, D. (2004). Improved simulation of stabilizer circuits. Physical Review A - Atomic, Molecular, and Optical Physics, 70(5 A).

[12] Spekkens, R. W. (2008). Negativity and contextuality are equivalent notions of nonclassicality. Physical Review Letters, 101(2).

[13] Aharonov, D., Ben-Or, M. (1999). Fault-Tolerant Quantum Computation With Constant Error. arXiv:quant-ph/9906129

[14] Zeng, B., Cross, A., Chuang, I. L. (2011). Transversality versus universality for additive quantum codes. IEEE Transactions on Information Theory, 57(9), 6272-6284.

[15] Paternostro, M., Kim, M. S., Palma, G. M. (2003). Non-local quantum gates: a cavityquantumelectrodynamics implementation. arXiv:quant-ph/0207043

[16] Ozols, M. (2008). Clifford group Pauli matrices Clifford group. arXiv, 6-9.

[17] Van den Nest, M. (2010). Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond. Quantum Information and Computation, vol. 10(3).

[18] Eastin, B., Knill, E. (2009). Restrictions on transversal encoded quantum gate sets. Physical Review Letters, 102(11).

[19] Wallman, J., Bartlett, S. (2012). Non-Negative Subtheories and Quasiprobability Representations of Qubits. Phys. Rev. A 85, 062121