Magic-State Distillation in All Prime Dimensions Using Quantum Reed-Muller Codes

Earl T. Campbell,[1,*] Hussain Anwar,[2] and Dan E. Browne[2]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*

(Received 29 May 2012; revised manuscript received 19 July 2012; published 27 December 2012)

We propose families of protocols for magic-state distillation—important components of fault-tolerance schemes—for systems of odd prime dimension. Our protocols utilize quantum Reed-Muller codes with transversal non-Clifford gates. We find that, in higher dimensions, small and effective codes can be used that have no direct analogue in qubit (two-dimensional) systems. We present several concrete protocols, including schemes for three-dimensional (qutrit) and five-dimensional (ququint) systems. The five-dimensional protocol is, by many measures, the best magic-state-distillation scheme yet discovered. It excels both in terms of error threshold with respect to depolarizing noise (36.3%) and the efficiency measure known as yield, where, for a large region of parameters, it outperforms its qubit counterpart by many orders of magnitude.

DOI: 10.1103/PhysRevX.2.041021 Subject Areas: Quantum Information

I. INTRODUCTION

The central challenge of implementing scalable quantum computing is to protect quantum systems against noise and decoherence while retaining the capacity to perform computation. Quantum error correction and fault-tolerant techniques provide a solution to this problem, and a variety of constructions for fault-tolerant quantum computation have been proposed [1–4]. In all these schemes, a delicate balance must be maintained between coherently manipulating the encoded system while preserving the protected subspace and prohibiting the proliferation of errors. For example, for schemes built on stabilizer codes [5], transversal gates have the desired properties, while, in topological systems, topologically protected braiding operations [2] provide the logical gates. While much work in quantum computation has focused on qubits (two-level systems), it is known that, for any prime d, effective codes exist for storing d-level quantum systems [5–7]. Thus, qudit systems are also candidates for scalable fault-tolerant quantum computation.

In many approaches, the protected unitary gates are a subset of the so-called Clifford group. It is known that the stabilizer operations (comprising Clifford unitaries as well as preparation and measurements in the computational basis) can be classically simulated efficiently [5,6,8] and that, on their own, they are not universal for quantum computation. Furthermore, several theorems have shown [9–12] that, in general, there is a tension between providing protection against generic noise and achieving universal quantum computing.

Despite these obstacles, fault-tolerant universal quantum computing is possible [1]. One particularly successful approach, known as state injection, is to achieve universality by augmenting the fault-tolerant operations with a supply of many copies of a suitable ancillary resource state. While methods for the direct preparation of sufficiently noise-free protected resource states have been proposed [1], a particularly elegant solution can be provided by distillation techniques, outlined in Fig. 1, where many noisy copies of a resource state can be distilled to arbitrary fidelity by using only error-protected operations while preserving the error threshold of the model.

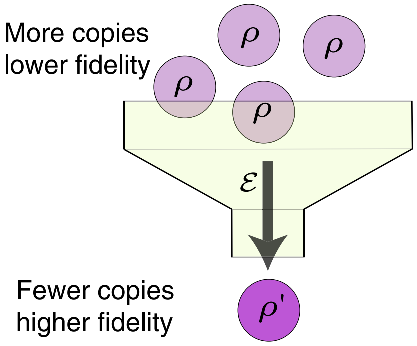

FIG. 1. An outline of a single round of magic-state-distillation protocol. Within many architectures of fault-tolerant quantum computing, a large proportion of the device is committed to these magic-state factories. Each attempt uses n copies of a state ρ and when successful outputs a state $\rho' \propto \mathcal{E}(\rho^{\otimes n})$. For n given successful attempts, the output states are used as inputs into the next iteration. Within the magic-state model, the completely positive map, $\mathcal{E}$, is composed of a sequence of Clifford unitaries and Pauli measurements. This figure illustrates a protocol where $n = 4$, for example, the ququint, $d = 5$, protocol that we discuss throughout the article.

[*] earltcampbell@gmail.com

Here, we consider the typical but idealized case in which the available protected operations are perfect stabilizer operations. In our discussion, whenever we speak of a qubit, we mean a qubit encoded into either a topological system or a stabilizer code that provides protection of stabilizer operations. This idealized setup has become known as the magic-state model. It was first studied by Bravyi and Kitaev [13], who proposed two protocols for the distillation of high purity nonstabilizer states. The states produced by distillation are a suitable resource for subsequent state injection of a non-Clifford unitary and so are known as magic states. In parallel, Knill proposed the concept of a postselected quantum computer [4,14,15] that used state preparation protocols that appeared distinct from magic-state distillation but were later shown to be equivalent [16]. These techniques are key components of many fault-tolerance schemes, including, for example, the topological cluster-state scheme [3,17–22].

Additional protocols discovered later by Reichardt [16,23,24] increased the family of qubit states known to be distillable. Conversely, Campbell and Browne [25,26] showed that no finite iterative protocol could distill all mixed nonstabilizer states. Results of many other investigations have contributed to our understanding of the magic-state theory for qubit systems [27–30], and a five-qubit distillation protocol has been implemented in an NMR system [31].

The theory of higher-dimensional quantum computation [32–35], stabilizer operations, and error-correcting codes [36,37] is well known. However, higher-dimensional magic-state models have been largely neglected until recently. In anyonic systems, the dimensionality of the available stabilizer operations is determined by the underlying physics, and so with some physical systems we would have no choice but to work in the higher-dimensional model (see, e.g., [38,39]). Recent progress on this problem has centered on exploiting a discrete phase-space, or Wigner function, representation of quantum states [40–42]. Notably, Veitch *et al.* [42] showed that states with positive Wigner functions can never be used as resources for magic-state distillation. Although all stabilizer states have positive Wigner functions, there also exist undistillable nonstabilizer states and thus bound magic states.

The discovery of this theorem, which rules out the possibility of magic-state distillation in certain cases, took place without any known distillation protocols in higher dimensions. However, we have recently proposed a protocol for three-dimensional (qutrit) systems that uses a generalization of the five-qubit code [43]. We observed magic-state distillation there, but the error suppression was slower than in qubit protocols. Here, we present a family of protocols that distill magic states in any odd prime dimension and do so with a quadratic reduction in noise per iteration. Given that noise reduction, the protocols are competitive with (and in some cases outperform) the best previously known qubit protocols.

Our protocols exploit higher-dimensional quantum Reed-Muller codes [44] and so generalize the qubit protocol of Bravyi and Kitaev [13] that used a 15-qubit quantum Reed-Muller code. This 15-qubit code was, to our knowledge, first developed by Knill, Laflamme, and Zurek [45] and later further developed by Steane [46]. These quantum codes are constructed from classical Reed-Muller codes [47–52], which have played a pivotal role in classical coding theory. (Notably, the family of Reed-Muller codes includes the infamous Reed-Solomon code used for communication with the Voyager space probe and for data storage on compact disks.)

We begin with a formal description of the Clifford group and the magic-state model. This description allows us to state our main theorem, which is roughly that magic-state distillation is possible in higher dimensions. Next, we review some basic theory of quantum error correction and show what properties of error-correcting codes would enable us to build a protocol for magic-state distillation. This review sets the stage for constructing codes, the quantum Reed-Muller codes, that have the required properties. Next, we introduce some additional tools from classical coding theory that help to simplify our analysis for a uniform depolarizing noise model.

To quantify the performance of our protocols, we consider several metrics and evaluate them for systems of up to 19 dimensions. In some respects, systems of different dimensions are incomparable, as the engineering challenges vary for systems of different sizes. However, if no additional assumptions on the physics of the underlying architecture are given, our figures of merit are the most natural. With this disclaimer, our analysis indicates that, for qutrits and ququints (five-dimensional systems), our protocols perform well, compared both to qubit protocols and to protocols for systems of even higher dimensions. For these two protocols, we investigate their performance in more detail. We find that the qutrit protocol performs well but that the ququint outperforms all other known magic-state protocols in terms of both the degree of error on the initial state it can tolerate and the efficiency of the protocol. We show that the effectiveness of these protocols can be related to properties of the Clifford group. The startlingly good performance of these protocols makes higher-dimensional systems an enticing alternative to qubit systems.

Finally, we show how to perform state injection to convert the distilled magic states into non-Clifford gates. The addition of any non-Clifford unitary to the set of n-qudit Clifford gates gives a set of gates dense in $\mathrm{SU}(d^n)$ and thus approximately universal via the Solovay-Kitaev theorem. This fact is well known for qubits but is also true for general prime d, as follows from theorems proven by Nebe, Rains, and Sloane in their recent book [53]. (See Appendix D.)

II. STABILIZER OPERATIONS AND THE MAGIC-STATE MODEL

We are interested in d-dimensional quantum systems, or qudits, where d is an odd prime. The computational-basis states are labeled by $j \in \mathbb{F}_d$, where $\mathbb{F}_d$ denotes the finite field of d elements. For such systems, the so-called Pauli group, $\mathcal{P}_d$, is generated by

$$X = \sum_{j\in\mathbb{F}_d} |j \oplus 1\rangle\langle j|, \qquad Z = \sum_{j\in\mathbb{F}_d} \omega^j |j\rangle\langle j|, \tag{1}$$

where $\oplus$ is addition modulo d and $\omega = \exp(i2\pi/d)$. The conjugation relation, $XZ = \omega^{-1}ZX$, is easy to verify and used throughout. The Pauli group over n qudits, $\mathcal{P}_d^n$, is the n-fold tensor product of the single qudit Pauli group. Consider an Abelian subgroup of the Pauli group $\mathcal{S}$ that contains the identity but no other multiple of the identity, e.g., $\omega\mathbb{1} \notin \mathcal{S}$. Associated with this group is a physical subspace, called a stabilizer code, and a projector onto this subspace, $\Pi \propto \sum_{s\in\mathcal{S}} s$. We equate Π with the code and call $\mathcal{S}$ the "stabilizer group" of the code. When the code is one-dimensional, the projector describes a pure quantum state, which we call a pure stabilizer state. We will also follow common terminology and call any probabilistic ensemble of pure stabilizer states a stabilizer state, even when a unique stabilizer group describing the mixture does not exist.

The Clifford unitaries $\mathcal{C}_d^n$ are those that conjugate the Pauli group to itself, so $\mathcal{C}_d^n = \{U; U\mathcal{P}_d^n U^\dagger = \mathcal{P}_d^n, U \in U(d^m)\}$. The whole Pauli group is a subgroup of the Clifford group, $\mathcal{P}_d^n \subset \mathcal{C}_d^n$. Gottesman [36] introduced several other Clifford gates, including the single qudit gates,

$$P = \sum_j \omega^{j(j-1)/2} |j\rangle\langle j|, \qquad H = \Big(\sum_{j,k} \omega^{jk} |j\rangle\langle k|\Big)\Big/\sqrt{d}, \tag{2}$$

and the two-qudit gate, the SUM gate (a generalization of the controlled bit-flip gate),

$$SUM = \sum_j |j\rangle\langle j| \otimes X^j. \tag{3}$$

These gates have been shown to generate the whole Clifford group [54]. The magic-state model also allows the implementation of so-called Pauli measurements. For any given Pauli $U \in \mathcal{P}_d^n$, which we express as $U = \sum_{k=0}^{d-1} \omega^k U_k$, we allow for positive-operator valued measurements with elements $\{U_k\}$. It is commonplace, although a modest abuse of terminology, to speak of measuring the Pauli U.

For an n-qudit system, the space of possible density matrices is within the set of bounded operators, $B(\mathcal{H}^{d^n})$, acting on $\mathcal{H}^{d^n}$. For such a space, the set of physical stabilizer operations allowed in the magic-state model is captured by the following definition.

Definition 1. Consider a completely positive map $\mathcal{E}$: $B(\mathcal{H}^{d^{n_{\rm in}}}) \rightarrow B(\mathcal{H}^{d^{n_{\rm out}}})$. The map is a stabilizer operation if and only if it can be composed from the following elements:

(1) Clifford unitaries,
(2) measurements and subsequent projections on stabilizer subspaces,
(3) preparation of fresh ancilla in a stabilizer state,
(4) tracing out of unwanted qudits, and
(5) adaptive decision making based both on measurement outcomes and on random coin tosses.

The number of qudits that are output and input may differ, as is typically the case when magic-state distillation is performed.

III. REQUIRED PROPERTIES OF GATES AND CODES

Every code defines an iterative scheme for magic-state distillation. However, some codes are much more suitable than others, and their usefulness can often be inferred from abstract properties of the code. In particular, the 15-qubit Reed-Muller code exploited by Brayvi and Kitaev has a very special property. There exists a product operator, of the form $U^{\otimes n}$, that acts on the logical basis as a non-Clifford operator. Such a code is said to have *transversal* non-Clifford gates, and we will consider generalizations of the qubit Reed-Muller codes with this remarkable property.

The transversal non-Clifford gate of the 15-qubit code, the so-called $\pi/8$ gate denoted as $U_{\pi/8}$, has another additional interesting property: For all Pauli $P \in \mathcal{P}_2^n$, we have that $U_{\pi/8} P U_{\pi/8}^\dagger \in \mathcal{C}_2^n$. Gottesman and Chuang defined this set of gates as the third level of an infinite hierarchy of qubit gates [55]. The hierarchy generalizes easily to qudits.

Definition 2. The kth level of the Clifford hierarchy for n qudits is the set

$$\mathcal{C}_d^n(k) = \{U | \forall P \in \mathcal{P}_d^n, UPU^\dagger \in \mathcal{C}_d^n(k-1)\}, \tag{4}$$

where the first level is the Pauli group $\mathcal{C}_d^n(1) = \mathcal{P}_d^n$.

The hierarchy is defined recursively with the kth level as the set of unitaries that conjugate the Pauli operators to a unitary in the $(k-1)$th level. The first level is fixed as the Pauli group, and the second level is simply the Clifford group $\mathcal{C}_d^n(2) = \mathcal{C}_d^n$. Higher levels are sets without a group structure. The qudit gates of interest share these properties with the qubit $U_{\pi/8}$ gate and are defined as follows.

Definition 3. The set of gates $\mathcal{M}_d^m$ contains all M such that

(1) M is diagonal in the computational basis,
(2) $M^{d^m} = \mathbb{1}$,
(3) $M \in {\rm SU}(d)$, and
(4) $M \in \mathcal{C}_d^1(3)/\mathcal{C}_d^1(2)$.

We now outline the motivation for these criteria and remark that M can be remembered as short for *magic*. Conditions 1–3 will be directly related to the transversality of the gate for our quantum Reed-Muller codes. Furthermore, if we express the eigenvalues of M as $\exp(i2\lambda_j\pi/d^m)$, then condition 2 entails that λ_j are integers, and condition 3 is satisfied when $\sum_j \lambda_j = 0$. Condition 4 requires that, while M is a member of the second level of the Clifford hierarchy, it is not a member of the Clifford group itself. Therefore, we conclude that the operator

$$C_M = MXM^\dagger \tag{5}$$

is in the Clifford group but is not a Pauli operator. The eigenstates of C_M will be the attractors of our distillation protocols, which is why it is essential that C_M is a non-Pauli operator. Distillation would be possible without requiring that C_M be a Clifford operator, but demanding this property provides us with tools that improve the protocol's efficiency. We observe that these sets form their own hierarchy such that, for any $m < m'$, we have $\mathcal{M}_d^m \subset \mathcal{M}_d^{m'}$. This relation holds because, almost trivially, $M^{d^{m+1}} = (M^{d^m})^d = \mathbb{1}^d = \mathbb{1}$. We remark also that, if $M \in \mathcal{M}_d^m$, then $M^\dagger \in \mathcal{M}_d^m$, and we use this feature throughout.

For every such set that is not empty, we design protocols that distill eigenstates of C_M. However, we need to know whether such gates exist. In the qubit setting, the $\pi/8$-phase gate provides such a unitary for $m = 4$. However, for $m < 4$, it is easy to check that all qubit gates with the form required by conditions 1–3 of the above definition are Clifford unitaries and thus fail condition 4. Remarkably, for all odd prime dimensions $d \geq 3$, we can find such gates for $m = 2$, and, when $d \geq 5$, these gates exist for $m = 1$, as was first shown in Ref. [56]. Using tall brackets to denote binomial coefficients, we have the following theorem.

Theorem 1. For all odd primes d, there exists a gate M such that:

(1) for $d = 3$, we have $M \in \mathcal{M}_d^m$ for all $m \geq 2$, and
(2) for prime $d \geq 5$, we have $M \in \mathcal{M}_d^m$ for all $m \geq 1$.

One such gate is the following:

$$M = \sum_j \exp(i2\lambda_j\pi/d^m)|j\rangle\langle j|, \tag{6}$$

with

$$\lambda_j = d^{m-2}\left[d\binom{j}{3} - j\binom{d}{3} + \binom{d+1}{4}\right]. \tag{7}$$

We refer to this M as the canonical $\mathcal{M}_d$ gate.

In particular, the canonical $\mathcal{M}_d$ gate is associated with the non-Pauli Clifford unitary

$$C_M = MXM^\dagger \propto XP, \tag{8}$$

where P is the Clifford gate introduced earlier in Eq. (2). Clearly, a different M exists for every dimension d. For notational clarity, we suppress this d dependence. To find a gate with the desired properties, we are guided by analogy with the qubit case. The qubit gate $U_{\pi/8}$ satisfies $U_{\pi/8}XU_{\pi/8}^\dagger \propto XS$, where $S = (\mathbb{1} + iZ)/\sqrt{2}$, and in the higher dimensions the gate P often plays an analogous role to S. The above similarities suggested to us that Eq. (8) might yield an M consistent with Definition 3. Given this informed hypothesis, solving for M requires only basic algebra, as is shown in Appendix A. Comprehensive classification of families of gates with the desired properties has been derived by Howard and Vala [56] using tools from symplectic geometry. That odd prime dimensions can produce the desired gates with smaller m is no mere technicality; it has far-reaching benefits for the magic-state distillation in higher dimensions. We also remark that these gates, for $d = 3, 5$, are Clifford equivalent to those found in Ref. [41] to be the most robust to depolarizing noise before becoming stabilizer operations.

The eigenstates of C_M are nonstabilizer states, which we label $|M_k\rangle$. We note that $|M_k\rangle = M|+_k\rangle$, where $|+_k\rangle$ is an eigenstate of X with eigenvalue ω^k. We aim to use magic-state distillation to purify copies of $|M_0\rangle$ from noisy copies and in turn to use these copies for fault-tolerant state injection of the magic unitary M. We are now in the position to state our main result.

Theorem 2. Consider any $M \in \mathcal{M}_d^m$ for any odd prime d and any integer $m \geq 2$, or any odd prime $d \geq 5$ and $m \geq 1$. There exists a stabilizer operation, $\mathcal{E}$, that iteratively distills the magic state $|M_0\rangle$. The map $\mathcal{E}$ takes $n = d^m - 1$ copies of a qudit state ρ, where

$$\epsilon = 1 - \langle M_0|\rho|M_0\rangle. \tag{9}$$

With nonzero probability, the protocol outputs a state $\rho' \propto \mathcal{E}(\rho^{\otimes n})$ such that

$$\epsilon' = 1 - \langle M_0^\dagger|\rho'|M_0^\dagger\rangle. \tag{10}$$

There exists a $K > 0$ such that for all ϵ we have $\epsilon' \leq K\epsilon^2$. Consequently, there exists a threshold $\epsilon^* > 0$ such that, if $0 < \epsilon < \epsilon^*$, then $\epsilon' < \epsilon$.

Notice that, after a single iteration, using noisy $|M_0\rangle$ states as input, the protocol will output a noisy $|M_0^\dagger\rangle$ state. By performing an even number of iterations, a fixed state can be distilled. We call this phenomenon cycling, and in many cases it may be prevented by some Clifford unitary correction. However, cycling can be desirable, as it provides us with a mechanism for producing both $|M_0\rangle$ and $|M_0^\dagger\rangle$ states. The rate of error suppression is always quadratic, and so these results give the first better-than-linear error reductions in higher-dimensional systems.

The Clifford unitary C_M plays a practical role in several steps of our protocols. First, it is used for C_M

twirling, which is a process for converting input states into a canonical form. By randomly choosing an integer, $k = 1, \ldots, d$, and applying C_M^k, we twirl any quantum state into the $|M_k\rangle$ basis. Hence, all qudit states, ρ, can be twirled into a form that depends only on $d - 1$-independent parameters, such that

$$\frac{1}{d}\sum_{k\in\mathbb{F}_d} C_M^k \rho (C_M^k)^\dagger = \sum_k f_k |M_k\rangle\langle M_k|. \tag{11}$$

Our distillation protocols seek to increase the value of f_0. Later, we show that C_M is also used in our protocols for *Clifford correction*, which significantly increases the success probability, and as part of the final state injection.

IV. MAGIC-STATE-DISTILLATION PROTOCOLS

A. CSS codes

Calderbank, Shor, and Steane identified a special class of quantum codes, which in their honor are now known as CSS codes [57]. These codes have stabilizers generated by two subgroups, $\mathcal{S}_Z$ and $\mathcal{S}_X$, which contain only Z^k and X^k terms, respectively. Therefore, the code projector has the form $\Pi_{\mathcal{S}} = \Pi_{\mathcal{S}_X}\Pi_{\mathcal{S}_Z}$. All CSS codes can also be described by a pair of classical vector spaces, which correspond to $\mathcal{S}_Z$ and $\mathcal{S}_X$. If we have a vector $\mathbf{u} \in \mathbb{F}_d^n$ and a single qudit operator, U, then we define the n-qudit operator

$$U[\mathbf{u}] = \otimes_{k=1}^n U^{u_k}. \tag{12}$$

The kth element of the vector, $\mathbf{u}$, tells us what multiple of U acts on the kth qudit. It follows that, for every $s \in \mathcal{S}_Z$, we can find a $\mathbf{u}$ such that $s = Z[\mathbf{u}]$. In fact, $\mathcal{S}_Z = \{Z[\mathbf{u}]; \mathbf{u} \in \mathcal{L}_Z\}$, where $\mathcal{L}_Z$ is a linear vector space. The closure of the stabilizer group under multiplication is easily seen to directly correspond to closure of $\mathcal{L}_Z$ under additional modulo d. Similarly, we can find a linear code, $\mathcal{L}_X$ for $\mathcal{S}_X$. The whole stabilizer must be Abelian, and so for all $\mathbf{u} \in \mathcal{L}_X$ and $\mathbf{v} \in \mathcal{L}_Z$ we require $\langle \mathbf{u}, \mathbf{v}\rangle = \oplus_j u_j v_j = 0$. Furthermore, for any code, $\mathcal{L}$, we define the dual code $\mathcal{L}^\perp = \{\mathbf{u}; \langle \mathbf{u}, \mathbf{v}\rangle = 0, \forall \mathbf{v} \in \mathcal{L}\}$. In terms of duality, commutation inside the stabilizer equates to $\mathcal{L}_X \subset \mathcal{L}_Z^\perp$ and $\mathcal{L}_Z \subset \mathcal{L}_X^\perp$. The dimensionalities of the duals are related by $\mathrm{Dim}(\mathcal{L}^\perp) = n - \mathrm{Dim}(\mathcal{L})$, where n is the dimension of the vector field they inhabit, namely, $\mathbb{F}_d^n$. For a CSS code, $k = n - \mathrm{Dim}(\mathcal{L}_Z) - \mathrm{Dim}(\mathcal{L}_X)$ gives the number of logical qudits supported by the code Π.

Here, we are solely interested in stabilizer codes of only d dimensions, in other words, a single logical qudit. It is useful to specify a basis spanning the code, which we again do using Pauli operators Z_L and X_L. These operators are the so-called logical operators of the subspace, and they must commute with the code stabilizer. However, with respect to each other, the logical operators must conjugate in the same way as Z and X, such that $X_\mathrm{L} Z_\mathrm{L} = \omega^{-1} Z_\mathrm{L} X_\mathrm{L}$. It follows that there exists an orthonormal basis, $\{|j_\mathrm{L}\rangle\}$, of stabilizer states that obey $Z_\mathrm{L}|j_\mathrm{L}\rangle = \omega^j |j_\mathrm{L}\rangle$, $X_\mathrm{L}|j_\mathrm{L}\rangle = |j_\mathrm{L} \oplus 1\rangle$, and which we call the logical basis. In this basis, the code projector can be expressed as $\Pi = \sum_j |j_\mathrm{L}\rangle\langle j_\mathrm{L}|$. We also make use of the X basis that we denote as $|+_j\rangle$ for single qudits stabilized by $\omega^{-j}X$ and $|+_j^\mathrm{L}\rangle$ for logical encoded states stabilized by $\omega^{-j}X_\mathrm{L}$. Typically, such logical operators can also be expressed in terms of vectors, such as $X_\mathrm{L} = X[\mathbf{u}]$, where commutation of X_L with $\mathcal{S}_Z$ entails $\mathbf{u} \subset \mathcal{L}_Z^\perp$ and $\mathcal{L}_Z \subset \mathbf{u}^\perp$.

For this given vector description, a useful fact is that $\mathcal{L}_Z = (\mathrm{span}(\mathcal{L}_X, \mathbf{u}))^\perp$, where the span$(\ldots, \ldots)$ is the vector space generated by its arguments. Let us prove this fact by first observing that, since $\mathcal{L}_Z \subset \mathbf{u}^\perp$ and $\mathcal{L}_Z \subset \mathcal{L}_X^\perp$, we have that $\mathcal{L}_Z \subset (\mathrm{span}(\mathcal{L}_X, \mathbf{u}))^\perp$. That $\mathcal{L}_Z$ can be no smaller than this set follows from dimension counting; more precisely,

$$\begin{aligned}\mathrm{Dim}\{(\mathrm{span}(\mathcal{L}_X, \mathbf{u}))^\perp\} &= n - \mathrm{Dim}(\mathrm{span}(\mathcal{L}_X, \mathbf{u})) \\ &= n - \mathrm{Dim}(\mathcal{L}_X) - 1.\end{aligned}$$

Since we have a single logical qudit, $k = 1$, we also know that $\mathrm{Dim}(\mathcal{L}_Z) = n - \mathrm{Dim}(\mathcal{L}_X) - 1$. Since the dimensionalities match, the assertion is proven. Also taking $Z_\mathrm{L} = Z[\mathbf{v}]$ and noting $(\mathcal{L}^\perp)^\perp = \mathcal{L}$, we can deduce many such results for single qudit codes by similar reasoning:

$$\mathcal{L}_Z = (\mathrm{span}(\mathcal{L}_X, \mathbf{u}))^\perp, \tag{13}$$

$$\mathcal{L}_Z^\perp = \mathrm{span}(\mathcal{L}_X, \mathbf{u}), \tag{14}$$

$$\mathcal{L}_X = (\mathrm{span}(\mathcal{L}_Z, \mathbf{v}))^\perp, \tag{15}$$

$$\mathcal{L}_X^\perp = \mathrm{span}(\mathcal{L}_Z, \mathbf{v}). \tag{16}$$

We employ the above relations throughout.

The smallest unitary capable of nontrivially acting on the code gives the code's robustness to noise. For CCS codes, it suffices to consider the phase and the bit-flip noise separately. For an operator $U[\mathbf{u}]$, its "size" is measured by the Hamming weight, $|\mathbf{u}|_\mathrm{H} = \{\#x_j; x_j \neq 0\}$, in other words, the number of qudits on which the operator acts nontrivially. The robustness to phase noise is measured by the distance, $D_Z = \min\{|\mathbf{v}|_\mathrm{H}; Z[\mathbf{v}]\Pi = Z_\mathrm{L}\Pi\}$, and to bit-flip noise, the robustness is measured by $D_X = \min\{|\mathbf{v}|_\mathrm{H}; X[\mathbf{v}]\Pi = X_\mathrm{L}\Pi\}$. The overall distance of the code is $D = \min\{D_X, D_Z\}$. Finally, we remark that for any code there always exists a Clifford unitary that *decodes*, such that $UZ_\mathrm{L}U^\dagger = Z_1$ and $UX_\mathrm{L}U^\dagger = X_1$.

B. Suitable codes

We now define the broad class of quantum codes that we show can be used to distill these magic states.

Definition 4. An n-qudit stabilizer code, Π, is an $\mathcal{M}_d^m$-distillation code if all of the following conditions hold:

(1) All $M \in \mathcal{M}_d^m$ are transversal such that $M^{\otimes n}\Pi(M^{\otimes n})^\dagger = M_\mathrm{L}^\dagger \Pi M_\mathrm{L}$,

(2) the code has a distance of $D \geq 2$, and

(3) the code has logical Pauli operators $X_{\mathrm{L}} = X[\mathbf{1}]$ and $Z_{\mathrm{L}} = Z[(d-1)\mathbf{1}]$.

We have introduced the vector shorthand $\mathbf{1} = (1, 1, \ldots, 1)$. Notice that we require a special kind of transversality, such that the logical operator, $M_{\mathrm{L}}^{\dagger}$, is implemented by applying $M^{\otimes n}$. The need for complex transposition is explained later and will be seen to result in a cycling phenomenon in the distillation protocol.

Here, we show that all $\mathcal{M}_d^m$-distillation codes can be used to perform distillation for magic states of the form $|M_0\rangle = M|+_0\rangle$ for all $M \in \mathcal{M}_d^m$. Because of cycling, after a single iteration using noisy $|M_0\rangle$ states as input, the protocol will output a noisy $|M_0^{\dagger}\rangle$ state.

Theorem 3. For a given n-qudit $\mathcal{M}_d^m$-distillation code of distance D, the following condition holds: For all $M \in \mathcal{M}_d^m$, there exists a stabilizer operation, $\mathcal{E}$, that iteratively distills the magic state $|M_0\rangle$. The protocol takes as input n copies of a state, ρ, where

$$\epsilon = 1 - \langle M_0|\rho|M_0\rangle. \tag{17}$$

With nonzero probability, the protocol outputs a state $\rho' \propto \mathcal{E}(\rho^{\otimes n})$ such that

$$\epsilon' = 1 - \langle M_0^{\dagger}|\rho'|M_0^{\dagger}\rangle. \tag{18}$$

There exists a $K > 0$ such that for all ϵ we have $\epsilon' \leq K\epsilon^D$. Consequently, there exists a threshold $\epsilon^* > 0$ such that if $0 < \epsilon < \epsilon^*$ then $\epsilon' < \epsilon$.

Later, we show the existence of the required codes with $D = 2$, which will then entail Theorem 2. For now, we show how to proceed when given such a code.

C. The protocol

We prove the above key result constructively. For a given n-qudit $\mathcal{M}_d^m$-distillation code and any $M \in \mathcal{M}_d^m$, we can perform the following iterative magic-state-distillation protocol.

(1) Take n copies of the state ρ and C_M twirl them.

(2) Measure generators of the phase stabilizer $\mathcal{S}_Z$.

(3) Accept all outcomes but perform a Clifford correction operator $C_M[\mathbf{w}]$ tuned to outcomes.

(4) Measure generators of the bit-flip stabilizer $\mathcal{S}_X$.

(5) Postselect on all "+1" measurement outcomes.

(6) Decode the encoded qudit to a single qudit.

(7) Use the output labeled ρ' as input in the next iteration.

When iterating the protocol, on the odd iterations we must replace C_M by $C_M^{\dagger}$ to account for cycling. We have not yet defined the exact setting of $C_M[\mathbf{w}]$ but will come to this in due time. For simplicity, though, we begin with assuming that step 2 generates all $+1$ measurement outcomes for which $C_M[\mathbf{w}] = \mathbb{1}$. We explain later how the Clifford correction in step 3 increases the success probability.

After C_M twirling the n copies, we have a state

$$\rho^{\otimes n} = \sum_{\mathbf{v} \in \mathbb{F}_d^n} \alpha_{\mathbf{v}} |M_{\mathbf{v}}\rangle\langle M_{\mathbf{v}}|, \tag{19}$$

where

$$|M_{\mathbf{v}}\rangle = |M_{v_1}\rangle|M_{v_2}\rangle \cdots |M_{v_n}\rangle \tag{20}$$

and

$$\alpha_{\mathbf{v}} = \prod_{k \in \mathbb{F}_d} f_k^{\mathrm{wt}_k(\mathbf{v})}, \tag{21}$$

where $\mathrm{wt}_k(\mathbf{v})$ is the k weight, the number of elements in $\mathbf{v}$ is equal to k, and $f_k = \langle M_k|\rho|M_k\rangle$. We note that

$$\rho^{\otimes n} = M_{\mathrm{L}}^{\dagger}\Big(\sum_{\mathbf{v} \in \mathbb{F}_d^n} \alpha_{\mathbf{v}} |+_{\mathbf{v}}\rangle\langle +_{\mathbf{v}}|\Big) M_{\mathrm{L}}, \tag{22}$$

where $M_{\mathrm{L}}^{\dagger} = M^{\otimes n}$. On a successful projection onto the code subspace, we have

$$\Pi\rho^{\otimes n}\Pi = M_{\mathrm{L}}^{\dagger}\Big(\sum_{\mathbf{v} \in \mathbb{F}_d^n} \alpha_{\mathbf{v}} \Pi|+_{\mathbf{v}}\rangle\langle +_{\mathbf{v}}|\Pi\Big) M_{\mathrm{L}}, \tag{23}$$

as the projector commutes with M_{L}. We need to determine the effect of each term $\Pi|+_{\mathbf{v}}\rangle$, which we find to be

$$\Pi|+_{\mathbf{v}}\rangle = 0; \quad \forall\ \mathbf{v} \notin \mathcal{L}_X^{\perp}; \tag{24}$$

$$\begin{aligned} \Pi|+_{\mathbf{v}}\rangle = \sqrt{c}\,|+_j^{\mathrm{L}}\rangle; \quad &\forall\ \mathbf{v} \oplus j\mathbf{1} = \mathbf{w}, \\ &\text{such that } \mathbf{w} \in \mathcal{L}_Z. \end{aligned} \tag{25}$$

The first equation covers all $\mathbf{v} \notin \mathcal{L}_X^{\perp}$, and the second equation covers all $\mathbf{v} \in \mathrm{span}(\mathcal{L}_Z, \mathbb{1})$. By virtue of Eq. (16), we know that $\mathcal{L}_X^{\perp} = \mathrm{span}(\mathcal{L}_Z, \mathbf{1})$, and so these equations account for all possible $\mathbf{v}$. The constant c gives the probability of this projection when the initial state is pure:

$$c = \mathrm{tr}(\Pi|+_0\rangle\langle +_0|^{\otimes n}). \tag{26}$$

Furthermore, $|+\rangle^{\otimes n}$ is an eigenstate of $\Pi_{\mathcal{S}_X}$, and so this randomness can be completely attributed to the Z-stabilizer measurements, which can be made deterministic by Clifford correction. Equations (24) and (25) follow directly from properties of error-correcting codes; for completeness, more details are given in Appendix B.

In summary, the transversality of M allows us to consider the distillation of magic states $|M_0\rangle$ as equivalent to the simpler problem of distillation in the X basis. Combining these results, we have

$$\Pi\rho^{\otimes n}\Pi = cM_{\mathrm{L}}^{\dagger}\Big(\sum_{j \in \mathbb{F}_d} \sum_{\mathbf{v} \oplus j\mathbf{1} \in \mathcal{L}_Z} \alpha_{\mathbf{v}} |+_j^{\mathrm{L}}\rangle\langle +_j^{\mathrm{L}}|\Big) M_{\mathrm{L}}. \tag{27}$$

The output state is diagonal in the basis $M_{\mathrm{L}}^{\dagger}|+_j^{\mathrm{L}}\rangle$ rather than in the desired $M_{\mathrm{L}}|+_j^{\mathrm{L}}\rangle$. We reiterate that this cycling is not problematic, as an even number of iterations always brings us back to the initial basis. Decoding onto a single qudit and using $\mathcal{E}$ to denote the whole process, we have

$$\rho' \propto \mathcal{E}(\rho^{\otimes n}) = c \sum_{j\in\mathbb{F}_d} \sum_{\mathbf{v}\oplus j\mathbf{1}\in\mathcal{L}_Z} \alpha_{\mathbf{v}} |M_j^\dagger\rangle\langle M_j^\dagger|. \tag{28}$$

By expanding $\alpha_{\mathbf{v}}$, we obtain an iterative formula for $f'_k = \langle M_k|\rho'|M_k\rangle$, such that

$$f'_j = \frac{\sum_{\mathbf{v}\oplus j\mathbf{1}\in\mathcal{L}_Z} \prod_{k\in\mathbb{F}_d} f_k^{\mathrm{wt}_k(\mathbf{v})}}{P}, \tag{29}$$

which has been renormalized by dividing through by the success probability P. This probability equals the sum of the numerators, which is

$$P = \sum_{j\in\mathbb{F}_d} \sum_{\mathbf{v}\oplus j\mathbf{1}\in\mathcal{L}_Z} \prod_{k\in\mathbb{F}_d} f_k^{\mathrm{wt}_k(\mathbf{v})}. \tag{30}$$

The summation over all j, such that $\mathbf{v}\oplus j\mathbf{1}\in\mathcal{L}_Z$, is equivalent to a sum over all $\mathbf{v}\in \mathrm{span}(\mathcal{L}_Z, (d-1)\mathbf{1})$. Using the features of CSS codes [see Eq. (15)], we know that $\mathrm{span}(\mathcal{L}_Z, (d-1)\mathbf{1}) = \mathcal{L}_X^\perp$, and so

$$P = \sum_{\mathbf{v}\in\mathcal{L}_X^\perp} \prod_{k\in\mathbb{F}_d} f_k^{\mathrm{wt}_k(\mathbf{v})}. \tag{31}$$

Notice that we have dropped a factor of c from the success probability, which will be justified later by Clifford correction. Both the numerator and the denominator of f'_j are polynomials of degree n and can be calculated from the classical codes.

D. Analyzing the iterative formulas

Here, we consider some properties of the above iterative formulas. First, we consider a simple depolarizing noise model and give a Taylor series approximation. Next, we consider a completely general noise model and show the existence of a distillation threshold.

When the noise is depolarizing, and so $f_{j\neq 0} = \epsilon/(d-1)$ and $f_0 = 1-\epsilon$, the formula for the fidelity simplifies to

$$f'_0 = \frac{\sum_{\mathbf{v}\in\mathcal{L}_Z} f_0^{n-|\mathbf{v}|_{\mathrm{H}}} f_{j\neq 0}^{|\mathbf{v}|_{\mathrm{H}}}}{\sum_{\mathbf{v}\in\mathcal{L}_X^\perp} f_0^{n-|\mathbf{v}|_{\mathrm{H}}} f_{j\neq 0}^{|\mathbf{v}|_{\mathrm{H}}}}, \tag{32}$$

where $|\ldots|_{\mathrm{H}}$ is again the Hamming weight. The factors f_0^n appear on both the numerator and the denominator and so cancel. Making use of the shorthand

$$\mu = \frac{f_{j\neq 0}}{f_0} = \frac{\epsilon}{(d-1)(1-\epsilon)}, \tag{33}$$

we can further simplify the fidelity formula to

$$f'_0 = \frac{\sum_{\mathbf{v}\in\mathcal{L}_Z} \mu^{|\mathbf{v}|_{\mathrm{H}}}}{\sum_{\mathbf{v}\in\mathcal{L}_X^\perp} \mu^{|\mathbf{v}|_{\mathrm{H}}}}. \tag{34}$$

Such cases are easier to study, as they depend only on a single parameter and the simple Hamming weights. Indeed, we show later, in Sec. V E, that this simple form can be further simplified by leveraging some powerful techniques from classical coding theory. For now, we make some casual observations concerning quadratic error suppression.

Taylor expanding the numerator and denominator to second order, we have

$$f'_0 \sim \frac{1 + a\mu^D + O(\mu^{D+1})}{1 + b\mu^D + O(\mu^{D+1})}, \tag{35}$$

where a (b) is the number of weight d elements of $\mathcal{L}_Z$ ($\mathcal{L}_X^\perp$), respectively. Both $\mathcal{L}_Z$ and $\mathcal{L}_X^\perp$ contain a single-weight zero element, $\mathbf{v} = \mathbf{0} = (0, 0 \ldots 0)$. By definition, both contain no other elements with weights smaller than d. Further approximating the denominator and using $f'_0 = 1 - \epsilon'$ yields

$$\epsilon' \sim (b-a)\mu^D + O(\mu^{D+1}). \tag{36}$$

So, the suppression of errors is degree D as $\mu \sim \epsilon$. In particular, since $D \geq 2$, the error suppression is at least quadratic.

The depolarizing noise model is useful for illustrating the salient features of a distillation protocol. However, it is important to demonstrate error suppression and the existence of a threshold for all possible noise models. Again, we rescale the noise parameters to $\mu_k = f_k/f_0$, and so

$$f'_0 = \frac{\sum_{\mathbf{v}\in\mathcal{L}_Z} \prod_{k=1}^{d-1} \mu_k^{\mathrm{wt}_k(\mathbf{v})}}{\sum_{\mathbf{v}\in\mathcal{L}_X^\perp} \prod_{k=1}^{d-1} \mu_k^{\mathrm{wt}_k(\mathbf{v})}}. \tag{37}$$

Both $\mathcal{L}_Z$ and $\mathcal{L}_X^\perp$ contain $\mathbf{v} = \mathbf{0}$, for which $\mathrm{wt}_{k\neq 0}(\mathbf{v}) = 0$, and so both the numerator and denominator contain a term equal to 1. We make a very coarse lower bound on the numerator, which must be greater than 1 since all terms are positive. We wish to set the upper bound on the denominator less coarsely. First, we define $\mu = \max_{k\neq 0}\{\mu_k\}$ and use it to replace all other noise parameters in the denominator, yielding the inequality

$$f'_0 \geq \left(\sum_{\mathbf{v}\in\mathcal{L}_X^\perp} \mu^{|\mathbf{v}|_{\mathrm{H}}} \right)^{-1}. \tag{38}$$

Recall that $\mathbf{v} = \mathbf{0}$ contributes 1 to the summation, and all other terms have an upper bound of μ^D, where D is the distance of the code. Hence, we have

$$f'_0 \geq (1 + C\mu^D)^{-1}, \tag{39}$$

where C is the number of nontrivial terms, $C = |\mathcal{L}_X^\perp| - 1$. Clearly, for real x we have $1 \geq (1 - x^2)$, and so $1 \geq (1-x)(1+x)$ and the positivity of x entails $(1+x)^{-1} \geq (1-x)$. Using this result with $x = C\mu^D$ gives $f'_0 \geq 1 - C\mu^D$, and furthermore

$$\epsilon' \leq C\mu^D \leq C\left(\frac{\epsilon}{1-\epsilon}\right)^D, \tag{40}$$

where $\epsilon' = 1 - f'_0$. We assume without loss of generality that f_0 is larger than all other f_j, which allows us to bound $(1-\epsilon)^{-1} \leq d$, and so

$$\epsilon' \leq d^D C \epsilon^D. \tag{41}$$

The above inequality gives us a valid constant $K = d^D C$, as asserted in Theorems 2 and 3. Realization of the existence of some distillation threshold follows quickly. If we consider $\epsilon^* = K^{-(D-1)^{-1}}$, we find that, if $0 < \epsilon < \epsilon^*$, then $\epsilon' < \epsilon$. The above analysis is very general, but the corresponding bounds are far from tight and a much higher ϵ^* exists.

E. Clifford correction

So far, we have assumed that the Z-stabilizer measurements all yield the desired $+1$ outcome. Next, we consider the process of Clifford correction, as outlined by step 3 of our protocol introduced in Sec. IV C. This additional strategy significantly increases the success probability of each round, so much so that success is guaranteed in the limit of pure initial states. The general idea is that, for any measurement outcomes with the resulting projector $\Pi'_{\mathcal{S}_Z}$, there exists a Clifford $C_M[\mathbf{w}]$ such that $C_M[\mathbf{w}]\Pi'_{\mathcal{S}_Z} = \Pi_{\mathcal{S}_Z} C_M[\mathbf{w}]$. The key fact exploited is that, for a single qudit, $C_M Z = \omega^{-1} Z C_M$, and so, for many qudits, $C_M[\mathbf{w}]Z[\mathbf{v}] = \omega^{-\langle \mathbf{w},\mathbf{v}\rangle} Z[\mathbf{v}] C_M[\mathbf{w}]$. To proceed, we must specify the projector $\Pi'_{\mathcal{S}_Z}$. We begin by expressing the linear code as $\mathcal{L}_Z = \{G\mathbf{u} \colon \mathbf{u} \in \mathbb{F}_d^m\}$, where $m = \mathrm{Dim}(\mathcal{L}_Z)$ and G is an $m \times n$ matrix called the generator matrix of $\mathcal{L}_Z$. Each column of G gives an individual generator of $\mathcal{L}_Z$ and hence of $\mathcal{S}_Z$. When the measurement corresponding to the jth generator gives the outcome ω^{k_j}, the resulting projection is

$$\Pi'_{\mathcal{S}_Z} = \frac{1}{2^m} \sum_{\mathbf{u} \in \mathbb{F}_d^m} \omega^{\langle \mathbf{k},\mathbf{u}\rangle} Z[G\mathbf{u}]. \tag{42}$$

Conjugating with a Clifford correction $C_M[\mathbf{w}]$ yields

$$C_M[\mathbf{w}]\Pi'_{\mathcal{S}_Z} = \frac{1}{2^m} \sum_{\mathbf{u} \in F_d^m} \omega^{\langle \mathbf{k},\mathbf{u}\rangle - \langle \mathbf{w}, G\mathbf{u}\rangle} Z[G\mathbf{u}] C[\mathbf{w}], \tag{43}$$

and so the correction works when for all $\mathbf{u}$ we have $\langle \mathbf{k}, \mathbf{u}\rangle = \langle \mathbf{w}, G\mathbf{u}\rangle$ modulo d. We can always choose a canonical form for the generator matrix, such that $G = (\mathbb{1}_m | G')$, where the identity acts on the first m rows of G and G' labels the remainder of the matrix. For such a canonical generator matrix, we choose $\mathbf{w}$ to equal $\mathbf{w} = (k_1, k_2, \ldots, k_m, 0, 0, \ldots, 0)$, so it matches the measurement outcomes on the first m entries. This choice yields $\langle \mathbf{w}, G\mathbf{u}\rangle = \langle \mathbf{k}, \mathbf{u}\rangle$, and so Clifford correction achieves its goal.

V. REED-MULLER CODES

A. Some concrete examples

Our demonstration of magic-state distillation in higher dimensions has been conditional on the existence of $\mathcal{M}_d^m$-distillation codes, as specified in Definition 4. Before introducing a family of $\mathcal{M}_d^m$-distillation codes for all odd prime d, we give some concrete examples. We label the codes as $\mathcal{QRM}_d(m)$, where d is again the dimensionality and m dictates the size and transversality properties of the codes.

Definition 5. $\mathcal{QRM}_3(2)$ is a CSS code over $n = 8$ qudits of dimension 3. The $\mathcal{L}_X$ code is generated by

$$\begin{aligned} \mathbf{u}_1 &= (1, 2, 0, 1, 2, 0, 1, 2), \\ \mathbf{u}_2 &= (0, 0, 1, 1, 1, 2, 2, 2). \end{aligned} \tag{44}$$

Similarly, $\mathcal{L}_Z$ is the code generated by

$$\begin{aligned} \mathbf{v}_1 &= (1, 2, 0, 1, 2, 0, 1, 2), \\ \mathbf{v}_2 &= (0, 0, 1, 1, 1, 2, 2, 2), \\ \mathbf{v}_3 &= (0, 0, 1, 2, 0, 2, 1, 0), \\ \mathbf{v}_4 &= (1, 1, 0, 1, 1, 0, 1, 1), \\ \mathbf{v}_5 &= (0, 0, 1, 1, 1, 1, 1, 1). \end{aligned} \tag{45}$$

The logical operators are $Z_{\mathrm{L}} = Z[2\mathbf{1}]$ and $X_{\mathrm{L}} = X[\mathbf{1}]$.

For the above qutrit code, we find that it is transversal with respect to the canonical $\mathcal{M}_3$ non-Clifford gate, as in Theorem 1,

$$M = \begin{pmatrix} \tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \tau^{-1} \end{pmatrix}, \tag{46}$$

where $\tau = \exp(i2\pi/9)$.

In the introduction of suitable non-Clifford gates, Theorem 1 shows that, for odd primes greater than 3, it is sufficient to set $m = 1$ to find non-Clifford gates. Remarkably, this property implies that we can find even smaller codes with transversal non-Clifford gates. The smallest such code exists for $d = 5$, and as we shall see later it performs exceptionally well at magic-state distillation.

Definition 6. $\mathcal{QRM}_5(1)$ is a CSS code over $n = 4$ ququints of dimension 5. The $\mathcal{L}_X$ code is generated by

$$\mathbf{u}_1 = (1, 2, 3, 4). \tag{47}$$

Similarly, $\mathcal{L}_Z$ is the code generated by

$$\mathbf{v}_1 = (1, 2, 3, 4), \qquad \mathbf{v}_2 = (1, 4, 4, 1). \tag{48}$$

The logical operators are $Z_{\mathrm{L}} = Z[4\mathbf{1}]$ and $X_{\mathrm{L}} = X[\mathbf{1}]$.

For the above code, we find that it is transversal with respect to the canonical $\mathcal{M}_5$ non-Clifford gate,

$$M = \begin{pmatrix} \omega^3 & 0 & 0 & 0 & 0 \\ 0 & \omega & 0 & 0 & 0 \\ 0 & 0 & \omega^{-1} & 0 & 0 \\ 0 & 0 & 0 & \omega^{-2} & 0 \\ 0 & 0 & 0 & 0 & \omega^{-1} \end{pmatrix}, \tag{49}$$

where $\omega = \exp(i2\pi/d) = \exp(i2\pi/5)$. Notice how the eigenvalues are all powers of ω. For dimensions smaller than $d = 5$, any diagonal gate with phases that are multiples of ω is a Clifford gate rather than a non-Clifford gate, as desired. This property makes it possible in higher dimensions to find smaller codes with a transversal non-Clifford gate.

From this information, one can numerically verify that both codes are well defined and have the correct transversality properties. Transversality can be verified by calculating the effect of the non-Clifford gates on the logical-basis states. Over the following sections, we develop an analytic proof that these features are valid for a whole family of quantum codes. Further details of the performance of these codes are given later, but we hope that these examples help guide the reader through the general case.

B. Classical Reed-Muller codes

Here, we review d-ary generalizations of Reed-Muller codes [49–52] and derive the crucial properties we exploit later. Convention dictates that we denote Reed-Muller codes as $\mathcal{RM}_d(u, m)$, where d tells us the relevant field, u is the order of the code, and m determines the size of the code. Here, we explicitly use only Reed-Muller codes of first order, so $u = 1$. All Reed-Muller codes are defined by polynomials of a degree bounded by the order of the codes. For order 1 Reed-Muller codes, we must consider degree 1 polynomials, in other words, linear functions. The dual of a Reed-Muller code is another Reed-Muller code, although it may have a different order [49–52]. In this way, higher-order Reed-Muller codes do enter into our work. However, it is sufficient for us to define them in terms of duality. Ultimately, we use not these codes but their smaller shortened versions introduced in the next section. However, for pedagogical reasons, we first review the unshortened variants.

We begin with a review of linear maps. There are d^m linear maps from $\mathbb{F}_d^m$ onto $\mathbb{F}_d$. All such maps, $g_{\bar{\mathbf{u}}}$: $\mathbb{F}_d^m \longrightarrow \mathbb{F}_d$, can be labeled by vectors themselves, say, $\bar{\mathbf{u}} \in \mathbb{F}_d^m$, and then the function will evaluate to $g_{\bar{\mathbf{u}}}(\mathbf{a}) = \langle \bar{\mathbf{u}}, \mathbf{a} \rangle = \oplus_j \bar{u}_j a_j$, again modulo d. Next, we consider another mapping, U_d^m: $\mathbb{F}_d^m \longrightarrow \mathbb{F}_d^n$, where $n = d^m$, such that

$$U_d^m(\bar{\mathbf{u}}) = (\langle \bar{\mathbf{u}}, \mathbf{a}_0 \rangle, \langle \bar{\mathbf{u}}, \mathbf{a}_1 \rangle, \ldots, \langle \bar{\mathbf{u}}, \mathbf{a}_{n-1} \rangle), \quad (50)$$

where $\mathbf{a}_j$ is the base d representation of the natural number j. For example, with $d = 3$ and $m = 2$, we have the ordered set

$$\{\mathbf{a}_j\} = \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2)\}.$$

Hence, for $\bar{\mathbf{u}} = (0, 1)$, we have

$$U_3^2[\bar{\mathbf{u}}] = U_3^2[(0, 1)] = (0, 1, 2, 0, 1, 2, 0, 1, 2). \quad (51)$$

For any d and m (positive integers), the set $\mathcal{L} = \{\mathbf{u} = U_d^m(\bar{\mathbf{u}}); \bar{\mathbf{u}} \in \mathbb{F}_d^m\}$ is a linear vector space. Closure of the vector space under addition follows directly from the closure under addition of homogenous linear maps. The codes of interest are constructed by considering all affine functions, which are linear maps plus an additional constant c such that they map $\bar{\mathbf{u}}$ to $U_d^m(\bar{\mathbf{u}}) \oplus c\mathbf{1}$.

Definition 7. Unshortened Reed-Muller codes, $\mathcal{RM}_d(1, m)$, are classical linear codes on $\mathbb{F}_d^n$, where $n = d^m$, of dimension $m + 1$. They are the set of code words $\mathcal{RM}_d(1, m) = \{U_d^m(\bar{\mathbf{u}}) \oplus c\mathbf{1}\colon \bar{\mathbf{u}} \in \mathbb{F}_d^m, c \in \mathbb{F}_d\}$ defined in terms of affine functions.

Such codes have many exotic properties. Before investigating them, we introduce one more definition.

Definition 8. We say that a function Λ: $\mathbb{F}_d^n \longrightarrow \mathbb{Z}$ is a λ function if there exists a set of d integers $\{\lambda_0, \ldots, \lambda_{d-1}\}$ such that $\sum_{j \in \mathbb{F}_d} \lambda_j = 0$ and

$$\Lambda(\mathbf{v}) = \sum_{j=1}^{n} \lambda_{v_j}. \quad (52)$$

Note that λ functions are closely related to the non-Clifford gates introduced in Definition 3. Our main observation here is the following lemma.

Lemma 1. Given a λ function Λ and an unshortened code $\mathcal{RM}_d(1, m)$, all $\mathbf{v} \in \mathcal{RM}_d(1, m)$ satisfy $\Lambda(\mathbf{v}) = 0$ modulo d^m.

To prove the lemma, we first consider code words where $\bar{\mathbf{u}} = \mathbf{0}$, and so $\mathbf{v} = (c, c, c \ldots c)$; then,

$$\Lambda(\mathbf{v}) = d^m \lambda_c, \quad (53)$$

which vanishes modulo d^m. Let us now consider the code word for the unit vector, $\bar{\mathbf{u}} = (1, 0, 0, \ldots, 0)$, and $c = 0$. The corresponding code word has a repetitive structure, as in Eq. (51), where each element of $\mathbb{F}_d$ appears d^{m-1} times. Hence,

$$\Lambda(\mathbf{v}) = d^{m-1} \sum_{j=0}^{d-1} \lambda_j = 0, \quad (54)$$

since we required in the definition of a λ function that $\sum_{j=0}^{d-1} \lambda_j = 0$. The above argument looks tailored to code words for a unit vector $\bar{\mathbf{u}}$, but a similar argument holds for all code words with nontrivial $\bar{\mathbf{u}}$. More precisely, for any nontrivial $\bar{\mathbf{u}}$, there are d^{m-1} different linear maps that evaluate to each possible output. To prove this statement, consider that the family of linear maps is invariant under change of variables that preserve linearity. Hence, the family of functions can always be expressed in a basis such that $\bar{\mathbf{u}}$ *is* a unit vector. Furthermore, these code words have uniform multiplicity of every value $\mathbb{F}_d$, and so adding $c\mathbf{1}$ reorders only the elements and not the multiplicity with which they appear. This argument proves our lemma.

In summary, unshortened Reed-Muller codes have a huge amount of symmetry that they inherit from the families of affine and linear maps. However, they actually have too much symmetry for our purposes. We break just enough of that symmetry by shortening the code.

C. Shortened classical Reed-Muller codes

Given a code $\mathcal{L}$ over $\mathbb{F}_d^n$, the corresponding shortened code, denoted as $\mathcal{L}^*$, is over $\mathbb{F}_d^{n-1}$. It contains all the code words of $\mathcal{L}$ with 0 in the first position and that position deleted. The process of shortening is closely related to puncturing, where the first position is removed but all code words are kept. We can also give a self-contained definition of a shortened Reed-Muller code as follows.

Definition 9. Shortened Reed-Muller codes, $\mathcal{RM}_d^*(1,m)$, are classical linear codes on $\mathbb{F}_d^n$, where $n = d^m - 1$, of dimension m. They are the set of code words $\mathcal{RM}_d^*(1,m) = \{P_d^m(\bar{\mathbf{u}})\colon \bar{\mathbf{u}} \in \mathbb{F}_d\}$ defined in terms of linear maps.

Here, P_d^m is the same map as U_d^m, but omitting the first element. For example, the shortened version of Eq. (51) is

$$P_3^2[\bar{\mathbf{u}}] = P_3^2[(0,1)] = (1,2,0,1,2,0,1,2), \tag{55}$$

which is also one of the generators of the $\mathcal{L}_X$ code for the quantum code, $\mathcal{QRM}_3(2)$, reviewed earlier. Notice that the self-contained definition of the shortened Reed-Muller code makes use of only linear maps and not affine maps. In the unshortened code, we had a generator $\mathbf{1}$ that corresponded to the constant term in affine functions. However, when shortening a code, we keep only code words with zero in the first position, and so the $\mathbf{1}$ generator is dropped. For this reason, the dimension of the code drops by one: $\mathrm{Dim}[\mathcal{RM}_d^*(1,m)] = \mathrm{Dim}[\mathcal{RM}_d(1,m)] - 1$. Let us now consider the shortened analog of Lemma 1.

Lemma 2. Given a λ function Λ and a shortened code $\mathcal{RM}_d^*(1,m)$, all $\mathbf{v} \in \mathcal{RM}_d^*(1,m)$ satisfy $\Lambda(\mathbf{v} \oplus c\mathbf{1}) = -\lambda_c$ modulo d^m.

The proof follows quickly from Lemma 1. Given a $\mathbf{v} \in \mathcal{RM}_d^*(1,m)$, let us define

$$\begin{aligned}\mathbf{w} &= (0, v_1, v_2, \ldots, v_n) \oplus c\mathbf{1},\\ &= (c, v_1 \oplus c, v_2 \oplus c, \ldots, v_n \oplus n),\end{aligned} \tag{56}$$

where clearly $\mathbf{w}$ is a code word of the unshortened code $\mathcal{RM}_d(1,m)$. Furthermore, $\Lambda(\mathbf{w}) = \Lambda(\mathbf{v}) + \lambda_c$, as it has an extra term appended. However, Lemma 1 tells us that $\Lambda(\mathbf{w}) = 0$, and so $\Lambda(\mathbf{v}) = -\lambda_c$. We will soon see that Lemma 2 is intimately related to the transversality of quantum gates for an associated quantum code.

D. Quantum Reed-Muller codes

Here, we construct quantum codes from shortened Reed-Muller codes for general m and d.

Definition 10. $\mathcal{QRM}_d(m)$ with $m \geq 1$ is a quantum CSS code over $n = d^m - 1$ qudits of prime dimension d. The code space is defined by

(1) $\mathcal{L}_X = \mathcal{RM}_d^*(1,m)$,
(2) $\mathcal{L}_Z = (\mathrm{span}(\mathcal{L}_X, \mathbf{1}))^\perp$,
(3) $X_\mathrm{L} = X[\mathbf{1}]$, and
(4) $Z_\mathrm{L} = Z[(d-1)\mathbf{1}]$.

We could have equivalently specified $\mathcal{L}_Z$ as a higher-order Reed-Muller code, although the above definition is simpler. We first check that $\mathcal{QRM}_d(m)$ codes are indeed quantum codes. By construction, the stabilizer is Abelian, as $\mathcal{L}_Z \subset \mathcal{L}_X^\perp$. It is easy to check that the logical operators are well defined: that Z_L commutes with the stabilizer, X_L commutes with the stabilizer, and $X_\mathrm{L}Z_\mathrm{L} = \omega^{-1}Z_\mathrm{L}X_\mathrm{L}$. Now, our next main result can be concisely stated.

Theorem 4. $\mathcal{QRM}_d(m)$ quantum codes are $\mathcal{M}_d^m$-distillation codes of distance $D = 2$.

The main property we need to prove is transversality for all $M \in \mathcal{M}_d^m$. As with all CCS codes, we have that

$$|j_\mathrm{L}\rangle = \frac{1}{\sqrt{|\mathcal{L}_X|}} \sum_{\mathbf{v} \in \mathcal{L}_X} |\mathbf{v} \oplus j\mathbf{1}\rangle. \tag{57}$$

Acting on this logical state with $M^{\otimes n}$ gives

$$M^{\otimes n}|j_\mathrm{L}\rangle = \frac{1}{\sqrt{|\mathcal{L}_X|}} \sum_{\mathbf{v} \in \mathcal{L}_X} \exp\left(i\frac{2\pi}{d^m}\Lambda(\mathbf{v} \oplus j\mathbf{1})\right)|\mathbf{v} \oplus j\mathbf{1}\rangle, \tag{58}$$

where Λ is a λ function (recall Definition 8) using the integers $\{\lambda_j\}$ associated with the eigenvalues of the unitary M. Now we use our key Lemma 2 to conclude

$$\begin{aligned}M^{\otimes n}|j_\mathrm{L}\rangle &= \frac{1}{\sqrt{|\mathcal{L}_X|}} \sum_{\mathbf{v} \in \mathcal{L}_X} \exp(-2i\pi\lambda_j/d^m)|\mathbf{v} \oplus j\mathbf{1}\rangle\\ &= \exp(-2i\pi\lambda_j/d^m)|j_\mathrm{L}\rangle = M_\mathrm{L}^\dagger|j_\mathrm{L}\rangle,\end{aligned}$$

and so we can identify $M^{\otimes n}$ with $M_\mathrm{L}^\dagger$.

Proving a distance lower bound is straightforward, as distance 2 is the smallest nontrivial distance. The relevant distance is D_z, the smallest $|\mathbf{v}|_\mathrm{H}$ such that it produces a logical error $Z[\mathbf{v}]\Pi = Z_\mathrm{L}^j\Pi$. For such an operator, $\mathbf{v} \in \mathcal{L}_X^\perp$ but $\mathbf{v} \neq 0$, so the phase error commutes with the X stabilizer but is nontrivial. If such an operator existed with Hamming weight 1, it would entail that there exists a qudit on which $\mathcal{L}_X$ acts trivially, which there is not. That the distance is not greater than 2 follows from the discussion in the next section.

E. MacWilliams identities

We have introduced higher-dimensional Reed-Muller codes and shown that they have suitable transversality properties for magic-state distillation. Knowing the code stabilizer and using Eqs. (29) and (34), we can calculate the exact analytic formula for arbitrary noise. For $\mathcal{QRM}_3(2)$, the general noise problem is tractable because $\mathcal{L}_Z$ and $\mathcal{L}_X^\perp$ are quite small sets, but the size and complexity of these sets grows rapidly with d and m. This complexity is relevant because the fidelity after one iteration is calculated by summing over all elements in $\mathcal{L}_Z$ and $\mathcal{L}_X^\perp$.

By considering depolarizing noise, the problem is partially simplified by Eq. (34), which we restate here as

$$f_0' = \frac{W_{\mathcal{L}_Z}(\mu)}{W_{\mathcal{L}_X^\perp}(\mu)}, \tag{59}$$

where $W_{\mathcal{L}}(\mu)$ is known as a weight enumerator,

$$W_{\mathcal{L}}(\mu) = \sum_{v\in\mathcal{L}} \mu^{|\mathbf{v}|_{\mathrm{H}}}. \tag{60}$$

Weight enumerators have been extensively studied in classical coding theory [47]. In particular, a weight enumerator for a code $\mathcal{L}$ can be related to the weight enumerator for the dual code $\mathcal{L}^\perp$ by the MacWilliams identity [47],

$$W_{\mathcal{L}^\perp}(\mu) = d^{-\mathrm{Dim}(\mathcal{L})}[1+(d-1)\mu]^n W_{\mathcal{L}}(\tilde{\mu}),$$

where we use the shorthand

$$\tilde{\mu} = \frac{1-\mu}{1+(d-1)\mu}. \tag{61}$$

Using $\mathcal{L}_Z = (\mathrm{span}(\mathcal{L}_X, \mathbf{1}))^\perp = (\mathcal{L}_X')^\perp$ [see Eq. (13)] and the MacWilliams identity, we have

$$f_0' = \frac{W_{\mathcal{L}_X'}(\tilde{\mu})}{dW_{\mathcal{L}_X}(\tilde{\mu})}. \tag{62}$$

The codes $\mathcal{L}_X$ and $\mathcal{L}_X'$ are much smaller and simpler than their duals, and so the MacWilliams identity has proven extremely helpful. Indeed, for Reed-Muller codes, we can find a closed form for these enumerators. When $\mathcal{L}_X = \mathcal{RM}_d(1,m)$, we have

$$W_{\mathcal{L}_X}(\tilde{\mu}) = 1 + (d^m - 1)\tilde{\mu}^{(d^m - d^{m-1})} \tag{63}$$

and

$$\begin{aligned} W_{\mathcal{L}_X'}(\tilde{\mu}) = W_{\mathcal{L}_X}(\tilde{\mu}) + (d-1)[\tilde{\mu}^{(d^m-1)} \\ + (d^m-1)\tilde{\mu}^{(d^m-1-d^{m-1})}]. \end{aligned} \tag{64}$$

(See Appendix C for details.) Combining all these formulas and reverting back to the original variables ϵ gives a closed analytic form, which is manageable, albeit a bit long for reproducing here. Rather, we present the Taylor expansion to second order in ϵ:

$$\epsilon' = \frac{(d^m-1)(d-2)}{2(d-1)}\epsilon^2 + O[\epsilon^3]. \tag{65}$$

It is interesting that, for all protocols based on a quantum code $\mathcal{QRM}_d(m)$, we see quadratic error suppression for all odd prime d and all m; in contrast, the quantum Reed-Muller code used by Bravyi and Kitaev, $\mathcal{QRM}_2(4)$, obtained a cubic reduction, such that $\epsilon' \sim 35\epsilon^3$. Our analysis also describes the Bravyi-Kitaev protocol, the only difference being that in the qubit case we need $m \geq 4$, and so the above formula also holds for qubits. It is intriguing to observe that the factor $(d-2)$ appears above, and so the quadratic term vanishes only in the qubit case; thus, in higher dimensions, these Reed-Muller codes are only distance 2. This property is one of many curious differences between qubits and odd prime dimensions.

VI. PERFORMANCE OF PROTOCOLS

Here, we consider various aspects of the performance of our protocols. We begin by showing that our protocols yield magic states at a rate that scales only polynomially with the desired final error probability. We then use MacWilliams identities to analyze thresholds under depolarizing noise models for much larger codes. Next, we consider in more detail the performance of our protocol based on $\mathcal{QRM}_3(2)$ and $\mathcal{QRM}_5(1)$.

A. Yields

The overall performance of a protocol can be captured by its yield. Given some target error probability, we calculate the yield as the expected fraction of the initial copies that achieves the goal. By definition, for any protocol and any distillable state ρ, with error probability ϵ_{in}, there exists a number of rounds $N(\rho, \epsilon_{\mathrm{target}})$ required to achieve $\epsilon_{\mathrm{target}}$. If on the kth round of distillation the success probability is P_k, the yield is simply

$$Y(\rho, \epsilon_{\mathrm{target}}) = \prod_{k=1,\ldots,N}\left(\frac{P_k}{n}\right), \tag{66}$$

where n is again the number of copies used per iteration. We are interested in how the yield scales as $\epsilon_{\mathrm{target}}$ vanishes. The success probability is continuous in ϵ and approaches 1 as ϵ vanishes; thus, P_k approaches 1 as k increases. Therefore, for all $p < 1$, there exists a c such that for all $k > c$ we have $P_k > P_c = p$. Hence, we can set a lower bound on the yield such that

$$Y(\rho, \epsilon_{\mathrm{target}}) \geq C\left(\frac{P_c}{n}\right)^{N-c}, \tag{67}$$

where C is a constant overhead, independent of $\epsilon_{\mathrm{target}}$, that represents the yield for c iterations. Furthermore, after c iterations, the error probability is now ϵ_c. Next, we observe that for a single round we know $\epsilon' \leq K\epsilon^D$ for some K, and equivalently $K\epsilon' \leq (K\epsilon)^D$. Therefore, the error probability after N iterations, ϵ_N, satisfies $K\epsilon_N \leq (K\epsilon_c)^{D^{N-c}}$. Taking $K\epsilon_c < 1$ allows us to bound the number of iterations needed such that

$$N - c < \left\lceil \log_D\left(\frac{\log(\epsilon_{\mathrm{target}}^{-1}/K)}{\log(\epsilon_c^{-1}/K)}\right)\right\rceil. \tag{68}$$

For positive a and b, we have the identity $a^{\log_D(b)} = b^{\log_D(a)}$, which, combined with the above equations, necessarily entails that

$$Y(\rho, \epsilon_{\mathrm{target}}) \geq C\left(\frac{\log(\epsilon_{\mathrm{target}}^{-1}/K)}{\log(\epsilon_c^{-1}/K)}\right)^{\log_D(P_c/n)}. \tag{69}$$

TABLE I. The yield-scaling parameter, γ^*, for distillation by $\mathcal{QRM}_d(m)$, as governed by Eq. (71). The smaller the value of γ^*, the more resource efficient the protocol in the limit of many iterations. For qubit systems, the 10-to-2 protocol of Ref. [58] achieves $\gamma^* = \log_2(5) \sim 2.321\,93$, which is the best-known value for qubit protocols. Empty cells indicate that no non-Clifford gate exists for those parameters.

d	$m = 1$	$m = 2$	$m = 3$	$m = 4$
2				2.464 97
3		3	4.700 44	6.321 93
5	2	4.584 96	6.9542	9.2854
7	2.584 96	5.584 96	8.417 85	11.2288
11	3.321 93	6.906 89	10.3772	13.8376
13	3.584 96	7.392 32	11.1007	14.8017
17	4	8.169 93	12.2621	16.3498
19	4.169 93	8.491 85	12.7436	16.9917

TABLE II. The distillation threshold $\epsilon^*_{\rm dep}$ for depolarizing noise when distilled by $\mathcal{QRM}_d(m)$. We include the threshold for the Brayvi-Kitaev protocol using 15 qubits, which uses a quantum Reed-Muller code $\mathcal{QRM}_2(4)$. Empty cells indicate that no non-Clifford gates exist for those parameters.

d	$m = 1$	$m = 2$	$m = 3$	$m = 4$
2				0.141 48
3		0.211 001	0.065 776 4	0.021 456 4
5	0.363 122 6	0.061 471 8	0.011 921 3	0.002 369 86
7	0.232 259 9	0.029 186 5	0.004 098 51	0.000 584 079
11	0.134 106 6	0.011 183 5	0.001 009 07	0.000 091 671 7
13	0.110 614 8	0.007 901 56	0.000 604 487	0.000 046 479 5
17	0.081 875 3	0.004 546 55	0.000 266 565	0.000 015 677 3
19	0.072 453	0.003 620 63	0.000 190 054	0.000 010 001 4

With the shorthand $\gamma = -\log_D(P_c/n)$, which is positive, we have

$$Y(\rho, \epsilon_{\rm target}) \geq C \frac{\log(\epsilon_c^{-1}/K)^\gamma}{\log(\epsilon_{\rm target}^{-1}/K)^\gamma}. \tag{70}$$

The above expression for the yield decreases by a factor polynomial in $\epsilon^{-1}_{\rm target}$. Conversely, the expected resource cost of distillation is the inverse yield, and this cost increases only polynomially in $\epsilon^{-1}_{\rm target}$. The scaling is governed by the factor $\gamma = -\log_D(P_c/n)$, but P_c can be taken arbitrarily close to 1. That being so, the relevant scaling parameter is $\gamma^* = \log_D(n)$, and so

$$Y(\rho, \epsilon_{\rm target}) \sim O[\log(\epsilon^{-1}_{\rm target}/K)^{-\gamma^*}]. \tag{71}$$

For our protocols in odd prime dimension, we find that $D = 2$ and $n = d^m - 1$, so $\gamma^* = \log_2(d^m - 1)$, which we give in Table I.

Notice that the code $\mathcal{QRM}_5(1)$ achieves the best yield scaling of all quantum Reed-Muller codes. This accolade is retained by $\mathcal{QRM}_5(1)$, even if we compare it with all presently known magic-state-distillation protocols.

B. Depolarizing noise thresholds

For some values of d and m, we have used the exact expression for ϵ' to find the depolarizing noise threshold $\epsilon^*_{\rm dep}$, below which distillation occurs (see Table II). However, these values should not be confused with the absolute threshold ϵ^* that holds for all noise models and can be smaller. The threshold gets weaker for both increasing d and increasing m, as suggested by the above approximate formula for ϵ' [see Eq. (65)]. When we increase m, we increase the number of copies required per iteration but decrease the depolarizing noise threshold. Consequently, it is advantageous to use the smallest possible m such that $M \in \mathcal{M}_d^m$. The benefit of larger m is, rather, that a large set of states is distilled by the protocol.

If we also compare our protocols with the threshold of the Bravyi-Kitaev (BK) protocol for $d = 2$, the pattern of better thresholds for smaller dimensions no longer holds. We see that the best threshold we observe is for $\mathcal{QRM}_5(1)$, with a fairly high threshold also observed for $\mathcal{QRM}_3(2)$. There are many subtle differences in the Clifford group between odd and even dimensions, and here those differences work in our favor. In odd prime dimension, we can construct smaller codes with transversal non-Clifford gates. Our code $\mathcal{QRM}_5(1)$ uses four ququints covering a Hilbert space of dimension 5^4, which to our knowledge is the smallest nontrivial stabilizer code with a transversal non-Clifford gate. Furthermore, research to date indicates that smaller codes lend themselves to better thresholds. A plausible explanation is that larger codes allow more undetected errors. Most of these undetected errors will have a large Hamming weight; thus, while they are negligible for small ϵ, they will be damaging for the modest sized ϵ relevant for threshold calculations.

Concerning thresholds for qubit protocols, we have focused on the comparable protocol using quantum Reed-Muller codes. The qubit threshold can be slightly extended by using the seven-qubit Steane code [16] ($\epsilon^* = 0.146\,45$) or the five-qubit code [13] on a different class of magic states ($\epsilon^* = 0.1719$). Although they represent a slight improvement, both fall short of our qutrit and ququint thresholds and have much poorer yields.

Before proceeding, we will remark on our notation and terminology for quantifying depolarizing noise. Throughout, we have used $\epsilon = 1 - \langle M_0|\rho|M_0\rangle$ for the error probability. If a state suffers depolarizing noise, it has the form

$$\rho = \delta|M_0\rangle\langle M_0| + (1 - \delta)\mathbb{1}/d, \tag{72}$$

and in some parts of the literature δ is used to quantify noise. Relating these two distinct noise measures, we have

$$\epsilon_{\rm dep} = (d-1)\delta/d, \qquad (73)$$

and so a dependence on the dimensionality appears. In terms of δ, thresholds appear larger, with $\mathcal{QRM}_3(2)$ and $\mathcal{QRM}_3(2)$ having thresholds at $\delta = 0.317$ and $\delta = 0.453$, respectively. Some readers may find using δ to be more natural, as it may be related to the depolarizing noise rate of some unitary used to prepare the initial noisy magic states. However, when unitaries suffer depolarizing noise, the best strategy is not to simply apply the noisy unitary to $|+\rangle$. Rather, better thresholds can be achieved with noisy unitaries by using the noise dilution protocol of Howard and Vala [56]. Furthermore, the threshold boosts from noise dilution become more prominent for higher dimensions.

C. Performance of $\mathcal{QRM}_3(2)$

Here, we apply our methods to the three-dimensional case using $\mathcal{QRM}_3(2)$, as explicitly defined in Definition 5. In previous work [43], we proposed other protocols for the three-dimensional case, including a generalization of the five-qubit code to qutrits. While magic-state distillation was observed for this five-qutrit code, these previous studies showed only a linear suppression of noise, whereas here we observe a more rapid quadratic suppression with each iteration.

We take M to be the canonical $\mathcal{M}_3$ gate, as in Theorem 1 and Eq. (46). By C_M twirling, all single qudit quantum states are projected onto the diagonal in the $|M_k\rangle$ basis, such that $\rho = \sum_k f_k |M_k\rangle\langle M_k|$. When we wish to distil $|M_0\rangle$, the weights f_1 and f_2 represent different types of noise. The parameter region for which the protocol is attracted to a magic state is shown in Fig. 2. A more convenient parametrization is $f_1 = \epsilon\cos^2(\theta)$ and $f_2 = \epsilon\sin^2(\theta)$, as we are mainly interested in how the total noise reduces. Our techniques allow us to find an analytic solution for ϵ' after a single iteration of magic-state distillation with $\mathcal{QRM}_3(2)$. However, the expression is lengthy, so here we truncate to third order,

$$\epsilon' = \epsilon^2[3+\cos(4\theta)] + \epsilon^3[9-\cos(4\theta)] + O[\epsilon^4], \qquad (74)$$

which is quadratically reduced. In Fig. 3(a), we show the exact output error probability for the whole range of different noise models (different θ) and depolarizing noise ($\theta = \pi/2$). We find that a threshold of $\epsilon^* = 0.200\,15$ for general noise and $\epsilon^*_{\rm dep} = 0.211\,001$ for depolarizing noise (as cited earlier). That being so, for all θ, if $0 < \epsilon < \epsilon^*$, it follows that $\epsilon' < \epsilon$. We can also find a quadratic upper bound, such that, for all ϵ and θ, we have $\epsilon' \leq K\epsilon^2$ with $K = 5.03$. The value of K is found by considering the function $\epsilon'\epsilon^{-2}$ and numerically maximizing, so $K = \sup_{\epsilon,\theta}\{\epsilon'\epsilon^{-2}\}$.

The region of distillable states is actually slightly larger than the $\epsilon < \epsilon^*$ region, with a greater noise tolerance for some values of θ. To find the whole distillable region, we resort to numerics and present the results as part of Fig. 2.

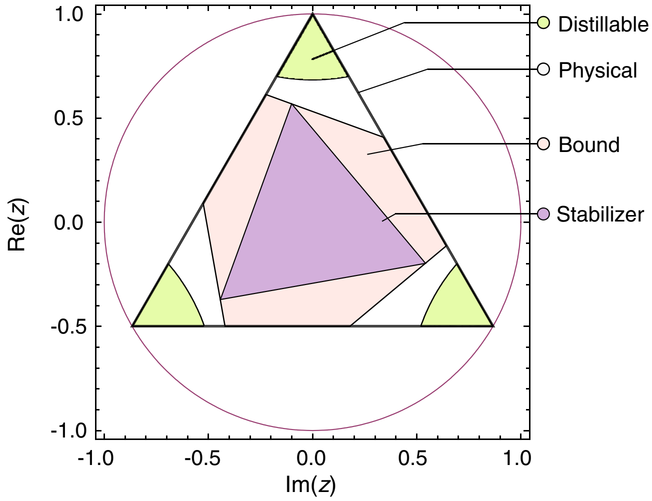

FIG. 2. The canonical C_M plane for a qutrit, $d = 3$, onto which any state can be projected by C_M twirling. Every quantum state is a point in the complex plane for the complex number $z_\rho = \mathrm{tr}(C_M\rho)$. The three pure magic states, $|M_k\rangle$, take values $z = 1, \omega, \omega^2$, which have $|z|^2 = 1$ and so lie on a circle in the plane. All *physical* states have $z = (1 - f_1 - f_2) + \omega f_1 + \omega^2 f_2$, and so lie in the convex hull of the pure magic states, forming a triangle of physical states. The *distillable* region of states can, by use of the $\mathcal{QRM}_3(2)$ protocol, be brought arbitrarily close to the nearest pure magic state. The *stabilizer* states are the convex hull over the set of points, z, taken for each of the pure stabilizer states. It is impossible to distill not only the stabilizer states but also the *bound* states, as demonstrated in Ref. [42]. Note that the rotational symmetry is to be expected, as the Pauli Z rotation performs a rotation in the C_M plane.

Several other important regions of the plane are also highlighted. We show the stabilizer states and bound magic states, which cannot be distilled by any stabilizer operation. Between these regions is a nonempty regime of ambiguous status, on which our protocol does not work and which is ruled out from distillability by any known theorem. Even in the simple qubit case, such puzzling regimes exist, and it has proven challenging to conclusively decide their status; see, for example, Refs. [25,26].

Also important is the success probability of distillation with $\mathcal{QRM}_3(2)$, which for all states satisfies $P \geq 1/9$ and for small ϵ is approximately

$$P = 1 - 8\epsilon + [31 + \cos(4\theta)]\epsilon^2 + O(\epsilon^3). \qquad (75)$$

Given these fairly high success probabilities and that we use only eight copies per iteration, this protocol is competitive in comparison to the Bravyi-Kitaev protocol that also used Reed-Muller codes. The BK protocol uses 15 copies per iteration and has $P \geq 1/16$, and for small ϵ it achieves $P = 1 - 15\epsilon + O(\epsilon^2)$. Our $\mathcal{QRM}_3(2)$ code requires fewer copies per iteration, but it would require more iterations to achieve the same error suppression as BK,

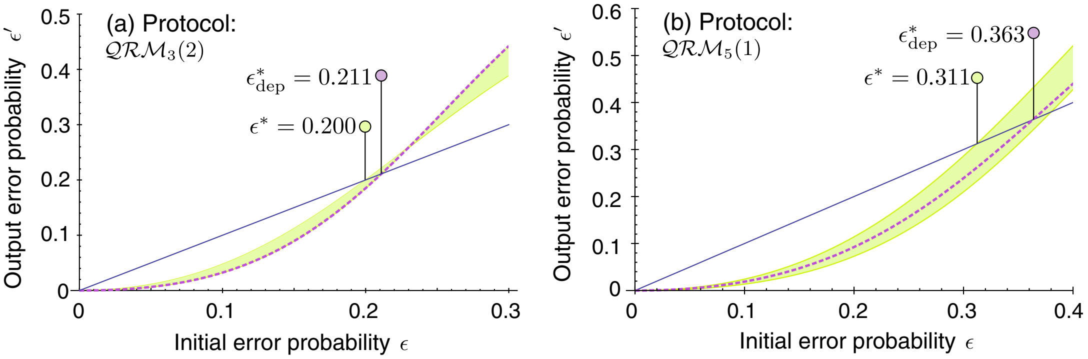

FIG. 3. The output error ϵ' against input error ϵ for (a) $\mathcal{QRM}_3(2)$ and (b) $\mathcal{QRM}_5(1)$. For a fixed ϵ, there are many different compatible states, and so there are many different possible output ϵ'. These output ϵ' are shown as a region rather than as a single curve. For the worst-case noise, we mark the threshold ϵ^*. The dashed line shows the specific instance of depolarizing noise, and the associated depolarizing threshold $\epsilon^*_{\rm dep}$ is also shown. The straight line is simply the "breakeven" line.

since BK has a cubic error suppression rather than just quadratic.

In Figs. 4(a) and 4(b), we consider the exact yield of our protocol, compared against BK, assuming depolarizing noise, so $\theta = \pi/4$. For small error probability $\epsilon_{\rm in} < 0.05$, the yield of our protocol $\mathcal{QRM}_3(2)$ is similar to BK. Both protocols give yields of the same order of magnitude, and whichever protocol is superior fluctuates with variation in required iterations. However, as the initial error probability $\epsilon_{\rm in}$ increases, the yield of $\mathcal{QRM}_3(2)$ exceeds that of BK by many orders of magnitude. The dominant effect here is that the yield of BK vanishes as we approach the threshold $\epsilon^*_{\rm BK} \sim 0.1415$, whereas our protocol can tolerate depolarization all the way up to $\epsilon^*_{\rm dep} \sim 0.211$.

The results of Sec. VI A also give us analytic tools for estimating yields. These tools show that for small $\epsilon_{\rm target}$ the yield of our protocol decreases as

$$Y \sim O[\log(\epsilon^{-1}_{\rm target}/5.03)^{-\gamma^*}], \tag{76}$$

where $\gamma^* = \log_2(8) = 3$. This scaling factor can be compared with the BK protocol, which achieves a similar scaling with $\gamma^* = \log_3(15) \sim 2.46$. We see that these protocols have similar scaling properties, but BK performs slightly better in the large $\epsilon^{-1}_{\rm target}$ limit. However, the numerical results reported in the previous paragraph show that finite size effects and a superior threshold often outweigh these asymptotic arguments.

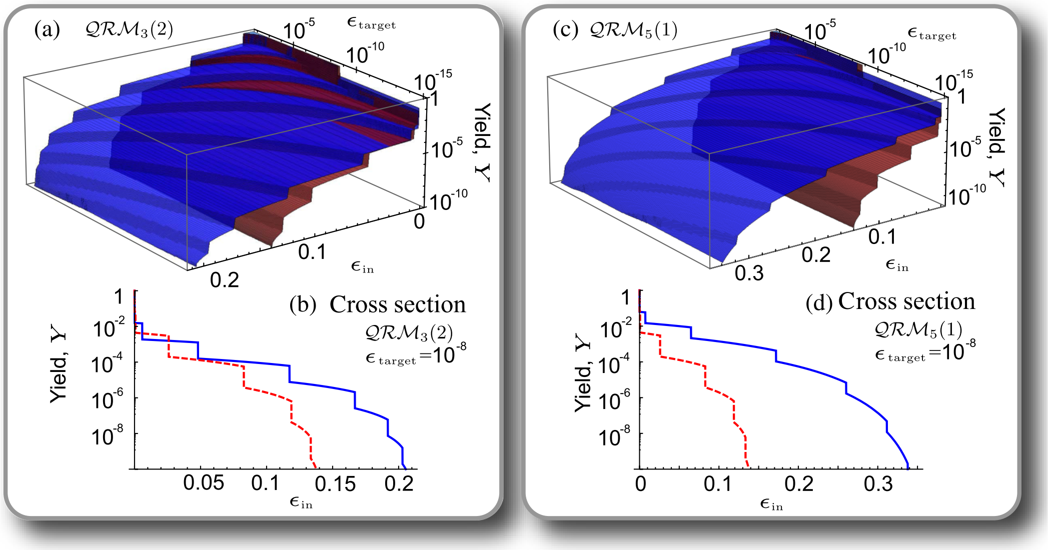

FIG. 4. The yield on a log scale of our protocols, $\mathcal{QRM}_3(2)$ and $\mathcal{QRM}_5(1)$ (blue), compared with the Bravyi-Kitaev $\mathcal{QRM}_2(4)$ (red) protocol. (a) and (c) are a function of initial error probabilities $\epsilon_{\rm in}$ and target error probabilities $\epsilon_{\rm target}$. For the qutrit and ququint states, the noise is depolarizing. (a) and (c) come with cross sections (b) and (d), respectively, with the target error probability held constant. The sudden changes in yield occur because of discrete changes in the number of iterations required.

D. Performance of $\mathcal{QRM}_5(1)$

Next, we apply our methods to the five-dimensional case using the code $\mathcal{QRM}_5(1)$, as explicitly defined in Definition 6. This protocol is the first ever applied to the problem of distilling magic states in five-dimensional systems. The code and associated protocol have many distinguishing features already mentioned: The code is the smallest known nontrivial code to have a transversal non-Clifford gate, has the largest noise threshold against depolarizing noise ($\epsilon^*_{\mathrm{dep}} = 0.363$), and has the best-known scaling in terms of expected yield (with $\gamma = 2$). All these features can be attributed to the fact that $d = 5$ is the smallest dimension where a diagonal non-Clifford gate exists with period d, allowing us to work with $m = 1$.

Again, we take M to be the canonical $\mathcal{M}_5$ gate, as in Theorem 1 and Eq. (49). The C_M-twirled states are parametrized by a fidelity, $f_0 = 1 - \epsilon$, and four independent noise parameters f_j for $j = 1, 2, 3, 4$. In Fig. 3(b), we show the range of different output error rates for all different types of noise and the depolarizing noise, which have thresholds of $\epsilon^* = 0.311\,95$ and $\epsilon^*_{\mathrm{dep}} = 0.363\,122$, respectively. We noted earlier that the $\mathcal{QRM}_5(1)$ possesses the best-known protection against depolarizing noise but also see here that its robustness against generic noise is also unrivaled.

Unfortunately, five-dimensional systems are quite complex. Even after twirling into the C_M plane, we cannot easily visually represent the whole distillability region as we did for the qutrit protocol. For this reason, we focus on the depolarized case with $f_0 = 1 - \epsilon$ and $f_{j\neq 0} = \epsilon/4$. After a successful implementation of one round, a depolarized state is output with

$$\epsilon' = \frac{\epsilon^2(96 - 160\epsilon + 75\epsilon^2)}{64 - 256\epsilon + 480\epsilon^2 - 400\epsilon^3 + 125\epsilon^4} \approx \frac{3\epsilon^2}{2} + \frac{7\epsilon^3}{2} \tag{77}$$

and occurs with probability

$$P = \frac{(1 - 2\epsilon)^4(64 - 256\epsilon + 480\epsilon^2 - 400\epsilon^3 + 125\epsilon^4)}{64(-1 + \epsilon)^4} \approx 1 - 8\epsilon + \frac{51\epsilon^2}{2}. \tag{78}$$

Based on these results, we expect the protocol to have an excellent yield. We numerically study the yield and again compare it against the qubit protocol $\mathcal{QRM}_2(4)$ or Bravyi and Kitaev; see Figs. 4(c) and 4(d). The numerics confirm that, across all parameter regimes, $\mathcal{QRM}_5(1)$ offers a significant resource savings of potentially many orders of magnitude. Magic-state distillation is typically the most resource intensive aspect of fault-tolerance schemes, and so high yield protocols are very desirable.

VII. STATE INJECTION AND UNIVERSAL QUANTUM COMPUTING

Protocols for qudit magic-state distillation are our main focus, but what happens after preparation of a highly purified magic state? Our ultimate goal is to simulate a non-Clifford group unitary via state injection. For the C_M-magic states of direct interest, we show the following.

Theorem 5. Consider any $M \in \mathcal{M}_d^m$ and any noisy magic state σ with $\epsilon = 1 - \langle M_0|\sigma|M_0\rangle$. There exists a trace-preserving stabilizer operation, $\mathcal{G}$, that deterministically implements state injection such that for all ρ

$$||\mathcal{G}(\sigma \otimes \rho) - M\rho M^\dagger||_1 \leq 2\epsilon, \tag{79}$$

where $||\ldots||_1$ is the trace norm $||A||_1 = \mathrm{tr}(\sqrt{AA^\dagger})$.

For perfect magic states, $\epsilon = 0$, this theorem entails that $\mathcal{G}(\rho_M \otimes \rho) = M\rho M^\dagger$. We first focus on the ideal case and later extend to noisy magic states.

For qubit systems, any magic state on the equator of the Bloch sphere may be exchanged for a unitary randomly selected from a pair of non-Clifford phase gates [13]. In a previous work, it was shown that a qutrit analog of the Bloch sphere equator [43] provides magic states that can be used for state injection of non-Clifford phase gates. Here, we review and generalize these ideas.

Definition 11. We say that a qudit quantum state $|\Theta\rangle$ is equatorial, or a phase state, if $\Theta \in \mathbb{R}^d$ and

$$|\Theta\rangle = \frac{1}{\sqrt{d}}\sum_{j=1}^{d} e^{i\Theta_j}|j\rangle. \tag{80}$$

It follows immediately that a $|M_0\rangle$ state is a phase state with $\Theta_j = 2\lambda_j\pi/d^m$. The essential feature of such states is that they are unbiased with respect to the computational basis, such that a Z measurement generates completely random outcomes. Taking an unknown state $|\psi\rangle$ and measuring $ZZ^\dagger$ on the pair $|\Theta\rangle|\psi\rangle$ also gives unbiased outcomes, and so no information is gained from $|\psi\rangle$. Denoting a general state as $|\psi\rangle = \sum_j c_j|j\rangle$, the result of a projection, Π_k, onto a subspace stabilized by $\omega^{-k}ZZ^\dagger$ yields

$$\Pi_k|\psi\rangle|\Theta\rangle \propto \sum_j c_j e^{i\Theta_{j\oplus k}}|j \oplus k\rangle|j\rangle. \tag{81}$$

We decode by performing a Clifford unitary such that $|j \oplus k\rangle|j\rangle \rightarrow |k\rangle|j\rangle$ and tracing out the first system. As promised, the result is a unitary transform, $|\psi\rangle \rightarrow U_k(\Theta)|\psi\rangle$, where

$$U_k(\Theta) = \sum_j e^{i\Theta_{j\oplus k}}|j\rangle\langle j|, \tag{82}$$

which can also be expressed as

$$U_k(\Theta) = (X^k)^\dagger U_0(\Theta)X^k, \tag{83}$$

where we note that

$$U_0(\Theta)|+_0\rangle = |\Theta\rangle. \tag{84}$$

The transformation is unitary but randomly selected from d different possibilities.

How do we simulate a deterministic unitary required for a computation? Herein, we consider unitary gates produced from a magic state, $|M_0\rangle$, such that

$$U_k = (X^k)^\dagger M X^k. \tag{85}$$

Again, we exploit the relationship between M and the Clifford unitary $C_M = MXM^\dagger$. Noting that $C_M^k = MX^kM^\dagger$, we express the unitary as

$$U_k = (X^k)^\dagger M X^k M^\dagger M \tag{86}$$

$$= (X^k)^\dagger C_M^k M. \tag{87}$$

Therefore, we can recover the desired M unitary by applying the inverse of Clifford unitary $(X^k)^\dagger C_M^k$.

We have established a deterministic stabilizer operation, such that $\mathcal{G}(|M_0\rangle\langle M_0| \otimes \rho) = M\rho M^\dagger$. We now relax our assumptions and allow the resource to be imperfect, so that $1 - \epsilon = \langle M_0|\sigma|M_0\rangle$. By C_M twirling, we can ensure that the state has the form $\sigma = (1-\epsilon)|M_0\rangle\langle M_0| + \epsilon\sigma'$, where $||\sigma'||_1 = 1$. Applying our map $\mathcal{G}$ to the noisy state gives

$$\mathcal{G}(\sigma \otimes \rho) = (1-\epsilon)M\rho M^\dagger + \epsilon\mathcal{G}(\sigma' \otimes \rho). \tag{88}$$

Subtracting $M\rho M^\dagger$ yields

$$\mathcal{G}(\sigma \otimes \rho) - M\rho M^\dagger = \epsilon\mathcal{G}(\sigma' \otimes \rho) - \epsilon M\rho M^\dagger. \tag{89}$$

Taking the trace norm and using the triangle inequality gives

$$||\mathcal{G}(\sigma \otimes \rho) - M\rho M^\dagger||_1 \\ \leq \epsilon||\mathcal{G}(\sigma' \otimes \rho)||_1 + \epsilon||M\rho M^\dagger||_1 = 2\epsilon.$$

The above is a rigorous treatment of the intuition that, if the magic state is almost perfect, then so too is the state injection.

The addition of non-Clifford M and $M^\dagger$ gates to our repertoire of unitaries generates a set dense in the special unitary group (see Refs. [53,59] and Appendix D). Furthermore, for every gate in this set, its inverse is also contained in the set. Thus, the Solovay-Kitaev algorithm can be applied to ensure an efficient approximation of any unitary. This argument also applies to the results of Ref. [43], where the qutrit Clifford group was supplemented by a non-Clifford unitary, but universality was only conjectured there.

VIII. DISCUSSION

We have generalized the idea of magic-state distillation using quantum Reed-Muller codes to all prime dimensions, enabling the preparation of highly purified nonstabilizer states for a given device capable of ideal stabilizer operations. By state injection, these magic states enable us to simulate universal quantum computation. While many aspects of the generalization were very analogous to the qubit case, there have also been some remarkable surprises. In odd prime dimension, the non-Clifford gates we gain are fundamentally different from the phase gates implemented by the Bravyi-Kitaev protocols. In particular, we find that for primes $d \geq 5$ there exist quantum Reed-Muller codes of only $d-1$ qudits that possess these non-Clifford gates as transversal gates, whereas $2^4 - 1 = 15$ qubits are needed for a similar construction.

To our knowledge, the ququint code ($d = 5$) using only four ququints is the smallest nontrivial stabilizer code with a transversal non-Clifford gate. Such a small code size translates into real practical gains, with the ququint protocol achieving better error-probability thresholds (see Sec. VI B) than any other known protocol with a polynomially scaling yield: $\epsilon^*_{\mathrm{dep}} = 0.363$ for depolarizing noise. Calculating the yield of the ququint protocol also shows that it is superior to all known qubit protocols, as is demonstrated by both numerics and analytic scaling arguments. For larger prime dimensions, $d > 5$, the thresholds and resource costs deteriorate with increasing dimension. It is not currently clear whether this deterioration is an inevitable problem with higher-dimensional systems or a peculiarity of our protocols. We also investigate in detail the performance of an eight-qutrit ($d = 3$) protocol, which, while not as effective as the ququint protocol, is still competitive against qubit protocols.

It is natural to question whether ϵ, as defined in Eq. (9), is a fair measure to use to compare noise thresholds in systems of different dimensions. It would be desirable to use a noise measure that is practically motivated based on noise processes that could occur in the lab. In Ref. [41], the depolarizing noise rate δ is employed, where δ measures the degree of depolarizing noise of state ρ from pure state $|\psi\rangle$ via $\rho = (1-\delta)|\psi\rangle\langle\psi| + (p/d)\mathbb{1}$. For a depolarizing noise model, δ is related to ϵ via $\epsilon = [(d-1)/d]\delta$. Quantifying error via δ penalizes higher-dimensional states, yet, even via this measure, the thresholds for the four-ququint code continue to significantly outperform their qubit counterparts.

Nevertheless, one can argue that δ is also an unfair method of comparison given the larger number of noise processes that contribute to depolarizing noise for higher-dimensional systems. Ultimately, for the context of magic-state distillation, the most relevant measure of comparison would be the yield at the fault-tolerance threshold. Unfortunately, at present, fault-tolerant quantum computation with higher-dimensional systems remains a little-explored research area, and thresholds comparable to, e.g., the schemes of Knill [4] or Harrington *et al.* [3], are unknown. We know of only one study of qudit fault-tolerance thresholds [60], and, while evidence was presented there that higher-dimensional systems may provide better thresholds than their binary counterparts, the analysis in that paper is limited. In particular, therefore, our results motivate further study of full fault-tolerance schemes based on ququint and qutrit components. It is possible that the enhanced performance in dimensions 3 and 5 seen in our magic-state-distillation protocols translates into better thresholds and resource costs for full fault-tolerance schemes based on qutrits and ququints.

Another application of our results is to models of computation where the fault-tolerant operations are a proper subgroup of the Clifford group. For instance, the qubit topological cluster states [3,17] cannot directly prepare Y eigenstates, but they can be distilled using magic-state distillation. In qudit generalizations of the topological cluster scheme, we anticipate that preparation of XZ eigenstates will not be topologically protected. While we have focused on the distillation of nonstabilizer states, our protocols also enable distillation of XZ stabilizer states.

Our understanding of the magic-state model is still in its infancy, despite many striking similarities to the more mature theory of entanglement. However, as we pointed out in our review in the Introduction, there has been a flurry of recent results on the qudit magic-state model. Numerous problems of a fundamental nature now present themselves as ripe for tackling. Inspired by entanglement theory, we might ask if qudit protocols exist for magic catalysis [61,62] or magic activation [61,63]. Furthermore, while all known protocols offer yields of magic states with arbitrarily small error probabilities, the yield vanishes as the target error vanishes. In contrast, in entanglement theory, the hashing protocol [64,65] and quantum polar-coding techniques [66] offer a method of distilling entanglement at a nonzero yield even for vanishing target error. Whether such a protocol could exist for magic-state distillation is an intriguing and wide-open question.

In the final stages of this research, we became aware of recent work that proposes a novel protocol [58] for qubit magic-state distillation. The protocol, which the authors of [58] call the 10-to-2 protocol, takes ten noisy magic states per iteration and outputs two magic states. This protocol is the first to output more than one magic state per iteration; it has the benefit of increasing its yield. Similar techniques could potentially also be used to design higher-dimensional protocols.

ACKNOWLEDGMENTS

We acknowledge the financial support of the EPSRC, the Leverhulme Trust, the BMBF (QuORep), and the EU (QESSENCE). D. E. B. was supported in part by the Perimeter Institute and would like to thank its members for their hospitality. We would like to thank Martin Kliesch, Daniel Gottesman, Peter Shor, Joe Fitzsimons, Debbie Leung, and Mark Howard for helpful discussions. In particular, we would like to thank Vadym Kliuchnikov for bringing Ref. [53] to our attention and for suggesting the argument given in Appendix. D.

APPENDIX A: THE CANONICAL M GATE

Here, we verify the assertions of Theorem 1 and show that the canonical M is a member of $\mathcal{M}_d^m$ for the asserted values of d and m. We begin by showing that

$$C_M = MXM^\dagger \propto XP. \tag{A1}$$

Left multiplying by $X^\dagger$ gives $X^\dagger MXM^\dagger \propto P$. The left-hand side is then

$$X^\dagger MXM^\dagger = \sum_j \exp[i2\pi(\lambda_{j\oplus 1} - \lambda_j)/d^m]|j\rangle\langle j|. \tag{A2}$$

The above equals P, up to a global phase, if, for all $0 \leq j \leq d-1$,

$$\lambda_{j\oplus 1} - \lambda_j = d^{m-1}\binom{j}{2} + c \tag{A3}$$

for some c. We first solve for the cases where $j \oplus 1 = j+1$, in other words, $j \neq d-1$. For this set of equations, we may use standard arithmetic and recurrence equation methods, and the general solution is

$$\lambda_j = d^{m-1}\binom{j}{3} + jc + \lambda_0 \tag{A4}$$

for all j, where c and λ_0 are integers to be determined. These integer variables will be fixed by demanding that Eq. (A3) with $j = d-1$ holds, and also that $\sum_j \lambda_j = 0$. First, let us impose the former condition and substitute Eq. (A4) into Eq. (A3) for $j = d-1$, to yield

$$\begin{aligned}\lambda_0 - \lambda_{d-1} &= \lambda_0 - \left[d^{m-1}\binom{d-1}{3} + j(d-1)c + \lambda_0\right]\\ &= d^{m-1}\binom{j-1}{2} + c.\end{aligned}$$

Solving this equation for c yields

$$c = -d^{m-2}\binom{d}{3}. \tag{A5}$$

For $m \geq 2$, inspection reveals that c is integer valued for all d. For $m = 1$, c is integer valued for all prime $d \geq 5$, which follows from the fact that, when $m = 1$, $c = -(d-1)(d-2)/6$. We use the fact that $6 = 3 \times 2$. Since $d \geq 5$ is a prime number not equal to three d, it is not divisible by 3; thus, either $(d-1)$ or $(d-2)$ must be divisible by 3. Since $d \geq 5$ is a prime number not equal to 2, then $(d-1)$ must be divisible by 2. Hence, the product $(d-1)(d-2)$ is divisible by 6 for all primes $d \geq 5$ and c is an integer for $m = 1$ and $d \geq 5$.

It remains to fix λ_0 by imposing that $\sum_j \lambda_j = 0$. Performing the summation and simplifying, we find that

$$\lambda_0 = d^{m-2}\binom{d+1}{4}. \tag{A6}$$

For $m \geq 2$, we see, by inspection, that λ_0 is integer valued for all d. For $m = 1$, λ_0 is integer valued for all prime $d \geq 5$, and the proof for this latter case is as follows. When $m = 1$, $\lambda_0 = (d+1)(d-1)(d-2)/24$. We observe that $24 = 3 \times 2 \times 4$. Since $d \geq 5$ is a prime number not equal to three d, it is not divisible by 3; thus, either $(d-1)$ or $(d-2)$ must be divisible by 3. Since d is an odd prime number, both $(d+1)$ and $(d-1)$ must be divisible by 2, and one of this pair must be divisible by 4. Hence, $(d+1)(d-1)(d-2)$ must be divisible by 24, and consequently λ_0 is an integer for $m = 1$ and $d \geq 5$.

Thus, the gate M, as defined in Theorem 1, satisfies all the requirements to be a member of $\mathcal{M}_d^m$. For $m = 1$ and $d = 3$, λ_j is not integer valued for all values of j, and so the above argument does not provide a member of $\mathcal{M}_3^1$. Indeed, for $d = 3$, it is easy to numerically search the sets of gates with integer λ_j and verify that none is non-Clifford and thus that $\mathcal{M}_3^1$ is empty.

APPENDIX B: PROJECTION ONTO LOGICAL SUBSPACE

Here, we present the reasoning that leads to Eqs. (24) and (25), which can be divided into three cases: a detected error, no error, and an undetected error.

When $\mathbf{v} \notin \mathcal{L}_X^\perp$, an error is present that is detected by the code, and so the state vanishes: $\Pi|+_\mathbf{v}\rangle = 0$. More precisely, we recall that $X|+_k\rangle = \omega^k|+_k\rangle$, and so more generally that $X[\mathbf{u}]|+_\mathbf{v}\rangle = \omega^{\langle \mathbf{v},\mathbf{u}\rangle}|+_\mathbf{v}\rangle$. Projecting onto the $+1$ eigenspace of all $X[\mathbf{u}] \in \mathcal{S}_X$ entails that the state will vanish unless $\langle \mathbf{v}, \mathbf{u}\rangle = 0$ for all $\mathbf{u} \in \mathcal{L}_X$. In other words, it is simply the requirement that $\mathbf{v}$ is in the dual of $\mathcal{L}_X$, which proves Eq. (24).

For the no-error instances, $\mathbf{v} \in \mathcal{L}_Z$, the state does not vanish under projection. Furthermore, since $| + \mathbf{v}\rangle = Z[\mathbf{v}]|+\rangle^{\otimes n}$ and $\Pi Z[\mathbf{v}] = \Pi$, we have $\Pi|+_\mathbf{v}\rangle = \Pi|+\rangle^{\otimes n}$, and so all such states must be projected onto the same logical state. Finally, we observe that $|+\rangle^{\otimes n}$ is stabilized by $X_\mathrm{L} = X^{\otimes n}$, and so $\Pi|+\rangle^{\otimes n} = \sqrt{c}|+_0^\mathrm{L}\rangle$.

All other possibilities correspond to undetected errors, resulting in a projection onto other logical states. In such cases, $\mathbf{v} \in \mathcal{L}_X^\perp$, and so there must exist a $j \in \mathbb{F}_d$ such that $\mathbf{w} = \mathbf{v} \oplus j\mathbf{1} \in \mathcal{L}_Z$. In terms of Pauli operators, we have $Z[\mathbf{w}] = Z[\mathbf{v}]Z[j\mathbf{1}]$, and so $Z[\mathbf{v}] = Z[\mathbf{w}]Z[(d-j)\mathbf{1}]$. Since the logical operator is $Z_\mathrm{L} = Z[(d-1)\mathbf{1}]$, it follows that $Z[\mathbf{v}] = Z[\mathbf{w}]Z_\mathrm{L}^j$. In terms of the quantum state, we have $|+_\mathbf{v}\rangle = Z[\mathbf{w}]Z_\mathrm{L}^j|+\rangle^{\otimes n}$, and so after projection $\Pi|+_\mathbf{v}\rangle = \sqrt{c}Z_\mathrm{L}^j|+_0^\mathrm{L}\rangle = \sqrt{c}|+_j^\mathrm{L}\rangle$.

APPENDIX C: WEIGHT ENUMERATORS

Here, we find the weight enumerators for the shortened Reed-Muller codes $\mathcal{L}_X = \mathcal{RM}_d^*(1, m)$ and $\mathcal{L}_X' = \mathrm{span}(\mathcal{L}_X, \mathbf{1})$, as given in Eqs. (63) and (64). Since $\mathcal{L}_X \subset \mathcal{L}_X'$, it is natural to start with $\mathcal{L}_X$ and then add the remaining terms.

First, $\mathcal{L}_X$ contains a zero vector $(0, 0, \ldots, 0)$ with zero Hamming weight. Second, all the remaining code words—there are $d^m - 1$ such code words—have $(d-1)$ zeros, i.e., have Hamming weight $n - (d-1) = d^m - d$. Thus, we have the weight enumerator

$$W_{\mathcal{L}_X}(x) = 1 + (d^m - 1)x^{(d^m - d)}. \tag{C1}$$

The enumerator for $\mathcal{L}_X'$ can be broken into d separate sums, since $\mathcal{L}_X' = \{\mathcal{L}_X, \mathcal{L}_X \oplus \mathbf{1}, \ldots, \mathcal{L}_X \oplus (d-1)\mathbf{1}\}$, and so

$$W_{\mathcal{L}_X'}(x) = \sum_{j=0}^{d-1} W_{\mathcal{L}_X \oplus j\mathbf{1}}(x) = W_{\mathcal{L}_X}(x) + \sum_{j=1}^{d-1} W_{\mathcal{L}_X \oplus j\mathbf{1}}(x).$$

For the rest of this argument, we focus on the $j \neq 0$ terms. First, each $j\mathbf{1}$, when added to the $(0, 0, \ldots, 0)$ vector, generates a code word of full Hamming weight $(n = d^m - 1)$. Second, each $j\mathbf{1}$, when added to any other code word of $\mathcal{L}_X$ [other than the $(0, 0, \ldots, 0)$ vector], results in a code word with d^{m-1} zeros, and so the Hamming weight is $n - d^{m-1} = d^m - 1 - d^{m-1}$. For each $\mathcal{L}_X \oplus j\mathbf{1}$, there are $d^m - 1$ such code words, and so

$$W_{\mathcal{L}_X \oplus j\mathbf{1}}(x) = x^{(d^m - 1)} + (d^m - 1)x^{(d^m - 1 - d^{m-1})}. \tag{C2}$$

For every $j \neq 0$, we obtain the same result. We have $d - 1$ such sums, and so

$$W_{\mathcal{L}_X'}(x) = W_{\mathcal{L}_X}(x) + (d-1)W_{\mathcal{L}_X \oplus \mathbf{1}}(x),$$

which expands into the formula given in the main text with $x = \tilde{\mu}$.

APPENDIX D: ADDING ANY NON-CLIFFORD GATE PROMOTES THE CLIFFORD GROUP TO A UNIVERSAL SET

We consider quantum circuits on n qudits of odd prime dimension d. We show here that the combination of two theorems of Nebe, Rains, and Sloane shows that the addition of any non-Clifford gate to the Clifford group generates a set of unitaries that is dense in $\mathrm{SU}(d^n)$. In Ref. [59], their Theorem 7.3 implies that any finite group that contains the Clifford group must be generated by the Clifford group and a gate proportional to the identity. Thus, the group H generated by the Clifford group and a non-Clifford unitary (not proportional to the identity) cannot be finite and must be of infinite order.

In Corollary 6.8.2 of Ref. [53], the authors show that any closed subgroup H that satisfies $\mathcal{C}_d^n \subset H \subset \mathrm{U}(d^n)$ must either have finite order (ignoring global phase factors) or be $\mathrm{SU}(d^n)$. Combining this corollary with the above theorem by Nebe, Rains, and Sloane, we conclude that the closure of the group generated by the Clifford group and any non-Clifford unitary (not proportional to the identity) is $\mathrm{SU}(d^n)$.

[1] P. W. Shor, in *Proceedings of the 37th Annual Symposium on the Foundations of Computer Science, 1996* (IEEE, New York, 1996), p. 56–65.

[2] A. Yu. Kitaev, *Fault-Tolerant Quantum Computation by Anyons*, Ann. Phys. (Amsterdam) **303**, 2 (2003).

[3] Robert Raussendorf and Jim Harrington, *Fault-Tolerant Quantum Computation with High Threshold in Two Dimensions*, Phys. Rev. Lett. **98**, 190504 (2007).

[4] E. Knill, *Fault-Tolerant Postselected Quantum Computation: Threshold Analysis*, arXiv:quant-ph/0404104.

[5] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Ph.D. thesis, California Institute of Technology, 1997, arXiv:quant-ph/9705052.

[6] D. Gottesman, *Theory of Fault-Tolerant Quantum Computation*, Phys. Rev. A **57**, 127 (1998).

[7] Dorit Aharonov and Michael Ben-Or, *Fault-Tolerant Quantum Computation with Constant Error Rate*, SIAM J. Comput. **38**, 1207 (2008).

[8] Maarten Van den Nest, *Efficient Classical Simulations of Quantum Fourier Transforms and Normalizer Circuits over Abelian Groups*, arXiv:1201.4867v1.

[9] Xie Chen, Hyeyoun Chung, Andrew W. Cross, Bei Zeng, and Isaac L. Chuang, *Subsystem Stabilizer Codes Cannot Have a Universal Set of Transversal Gates for Even One Encoded Qudit*, Phys. Rev. A **78**, 012353 (2008).

[10] Bei Zeng, *Transversality Versus Universality for Additive Quantum Codes*, IEEE Trans. Inf. Theory **57**, 6272 (2011).

[11] Bryan Eastin and Emanuel Knill, *Restrictions on Transversal Encoded Quantum Gate Sets*, Phys. Rev. Lett. **102**, 110502 (2009).

[12] Sergey Bravyi and Robert Koenig, *Classification of Topologically Protected Gates for Local Stabilizer Codes*, arXiv:1206.1609.

[13] Sergey Bravyi and Alexei Kitaev, *Universal Quantum Computation with Ideal Clifford Gates and Noisy Ancillas*, Phys. Rev. A **71**, 022316 (2005).

[14] E. Knill, *Fault-Tolerant Postselected Quantum Computation: Schemes*, arXiv:quant-ph/0402171.

[15] E. Knill, *Quantum Computing with Realistically Noisy Devices*, Nature (London) **434**, 39 (2005).

[16] Ben W. Reichardt, *Quantum Universality from Magic States Distillation Applied to CSS Codes*, Quantum Inf. Process. **4**, 251 (2005).

[17] Robert Raussendorf, Jim Harrington, and Kovid Goyal, *Topological Fault-Tolerance in Cluster State Quantum Computation*, New J. Phys. **9**, 199 (2007).

[18] Simon J. Devitt, Austin G. Fowler, Ashley M. Stephens, Andrew D. Greentree, Lloyd C. L. Hollenberg, William J. Munro, and Kae Nemoto, *Architectural Design for a Topological Cluster State Quantum Computer*, New J. Phys. **11**, 083032 (2009).

[19] Austin G. Fowler and Kovid Goyal, *Topological Cluster State Quantum Computing*, Quantum Inf. Comput. **9**, 0721 (2009).

[20] Sean D. Barrett and Thomas M. Stace, *Fault Tolerant Quantum Computation with Very High Threshold for Loss Errors*, Phys. Rev. Lett. **105**, 200502 (2010).

[21] Y. Li, S. D. Barrett, T. M. Stace, and S. C Benjamin, *Fault Tolerant Quantum Computation with Nondeterministic Gates*, Phys. Rev. Lett. **105**, 250502 (2010).

[22] Ying Li and Simon C. Benjamin, *High Threshold Distributed Quantum Computing with Three-Qubit Nodes*, arXiv:1204.0443.

[23] Ben Reichardt, *Error-Detection-Based Quantum Fault-Tolerance Threshold*, Algorithmica **55**, 517 (2009).

[24] Ben W. Reichardt, *Quantum Universality by Distilling Certain One- and Two-Qubit States with Stabilizer Operations*, Quantum Inf. Comput. **9**, 1030 (2009).

[25] Earl T. Campbell and Dan E. Browne, *Bound States for Magic State Distillation in Fault-Tolerant Quantum Computation*, Phys. Rev. Lett. **104**, 030503 (2010).

[26] Earl T. Campbell and Dan E. Browne, *On the Structure of Protocols for Magic State Distillation*, Lect. Notes Comput. Sci. **5906**, 20 (2009).

[27] M. B. Plenio and S. Virmani, *Upper Bounds on Fault Tolerance Thresholds of Noisy Clifford-Based Quantum Computers*, New J. Phys. **12**, 033012 (2010).

[28] Wim van Dam and Mark Howard, *Tight Noise Thresholds for Quantum Computation with Perfect Stabilizer Operations*, Phys. Rev. Lett. **103**, 170504 (2009).

[29] N. Ratanje and S. Virmani, *Generalized State Spaces and Nonlocality in Fault-Tolerant Quantum-Computing Schemes*, Phys. Rev. A **83**, 032309 (2011).

[30] Jonas T. Anderson, *On the Power of Reusable Magic States*, arXiv:1205.0289.

[31] Alexandre M. Souza, Jingfu Zhang, Colm A. Ryan, and Raymond Laflamme, *Experimental Magic State Distillation for Fault-Tolerant Quantum Computing*, Nat. Commun. **2**, 169 (2011).

[32] Stephen D. Bartlett, Hubert de Guise, and Barry C. Sanders, *Quantum Encodings in Spin Systems and Harmonic Oscillators*, Phys. Rev. A **65**, 052316 (2002).

[33] D. L. Zhou, B. Zeng, Z. Xu, and C. P. Sun, *Quantum Computation Based on d-Level Cluster State*, Phys. Rev. A **68**, 062303 (2003).

[34] Michael J. Bremner, Dave Bacon, and Michael A. Nielsen, *Simulating Hamiltonian Dynamics Using Many-Qudit Hamiltonians and Local Unitary Control*, Phys. Rev. A **71**, 052312 (2005).

[35] Michael A. Nielsen, Michael J. Bremner, Jennifer L. Dodd, Andrew M. Childs, and Christopher M. Dawson, *Universal Simulation of Hamiltonian Dynamics for Quantum Systems with Finite-Dimensional State Spaces*, Phys. Rev. A **66**, 022317 (2002).

[36] Daniel Gottesman, *Fault-Tolerant Quantum Computation with Higher-Dimensional Systems*, Chaos Solitons Fractals **10**, 1749 (1999).

[37] Niel de Beaudrap, *A Linearized Stabilizer Formalism for Systems of Finite Dimension*, arXiv:1102.3354.

[38] J. R. Wootton, V. Lahtinen, B. Doucot, and J. K. Pachos, *Engineering Non-Abelian Topological Memories from Abelian Lattice Models*, Ann. Phys. (Amsterdam) **326**, 2307 (2011).

[39] David J. Clarke, Jason Alicea, and Kirill Shtengel, *Exotic Non-Abelian Anyons from Conventional Fractional Quantum Hall States*, arXiv:1204.5479.

[40] D. Gross, *Hudson's Theorem for Finite-Dimensional Quantum Systems*, J. Math. Phys. (N.Y.) **47**, 122107 (2006).

[41] Wim van Dam and Mark Howard, *Noise Thresholds for Higher-Dimensional Systems Using the Discrete Wigner Function*, Phys. Rev. A **83**, 032310 (2011).

[42] Victor Veitch, Christopher Ferrie, and Joseph Emerson, *Negative Quasiprobability Representation Is a Necessary Resource for Magic State Distillation*, arXiv:1201.1256v3.

[43] Hussain Anwar, Earl T. Campbell, and Dan E. Browne, *Qutrit Magic State Distillation*, New J. Phys. **14**, 063006 (2012).

[44] P. Sarvepalli and A. Klappenecker, in *Proceedings of the IEEE International Symposium on Information Theory, Adelaide, Australia, 2005* (IEEE, New York, 2005), p. 1023–1027.

[45] E. Knill, R. Laflamme, and W. Zurek, *Threshold Accuracy for Quantum Computation*, arXiv:quant-ph/9610011.

[46] A. Steane, *Quantum Reed-Muller Codes*, IEEE Trans. Inf. Theory **45**, 1701 (1999).

[47] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1988).

[48] E.F. Assmus and J.D. Key, *Designs and Their Codes* (Cambridge University Press, Cambridge, England, 1994).

[49] T. Kasami, Shu Lin, and W. Peterson, *New Generalizations of the Reed-Muller Codes—I: Primitive Codes*, IEEE Trans. Inf. Theory **14**, 189 (1968).

[50] E. Weldon, Jr., *New Generalizations of the Reed-Muller Codes—II: Nonprimitive Codes*, IEEE Trans. Inf. Theory **14**, 199 (1968).

[51] P. Delsarte, J.M. Goethals, and F.J. Mac Williams, *On Generalized Reed-Muller Codes and Their Relatives*, Inf. Control **16**, 403 (1970).

[52] Petra Heijnen and Ruud Pellikaan, *Generalized Hamming Weights of q-ary Reed-Muller Codes*, IEEE Trans. Inf. Theory **44**, 181 (1998).

[53] Gabriele Nebe, E. M. Rains, and N. J. A. Sloane, *Self-Dual Codes and Invariant Theory* (Springer, Berlin, 2006).

[54] S. Clark, *Valence Bond Solid Formalism for d-Level One-Way Quantum Computation*, J. Phys. A **39**, 2701 (2006).

[55] Daniel Gottesman and Isaac L. Chuang, *Demonstrating the Viability of Universal Quantum Computation Using Teleportation and Single-Qubit Operations*, Nature (London) **402**, 390 (1999).

[56] Mark Howard and Jiri Vala, *Qudit Versions of the Qubit* $\pi/8$ *Gate*, Phys. Rev. A **86**, 022316 (2012).

[57] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).

[58] Adam M. Meier, Bryan Eastin, and Emanuel Knill, *Magic-State Distillation with the Four-Qubit Code*, arXiv:1204.4221.

[59] Gabriele Nebe, E.M. Rains, and N.J.A. Sloane, *The Invariants of the Clifford Groups*, Des. Codes Cryptogr. **24**, 99 (2001).

[60] Aparna Kanungo, Master's thesis, Texas A&M University, 2005.

[61] Earl T. Campbell, *Catalysis and Activation of Magic States in Fault-Tolerant Architectures*, Phys. Rev. A **83**, 032317 (2011).

[62] Daniel Jonathan and Martin B. Plenio, *Entanglement-Assisted Local Manipulation of Pure Quantum States*, Phys. Rev. Lett. **83**, 3566 (1999).

[63] Paweł Horodecki, Michał Horodecki, and Ryszard Horodecki, *Bound Entanglement Can Be Activated*, Phys. Rev. Lett. **82**, 1056 (1999).

[64] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters, *Mixed-State Entanglement and Quantum Error Correction*, Phys. Rev. A **54**, 3824 (1996).

[65] I. Devetak and A. Winter, *Distillation of Secret Key and Entanglement from Quantum States*, Proc. R. Soc. A **461**, 207 (2005).

[66] J.M. Renes, F. Dupuis, and R. Renner, *Efficient Polar Coding of Quantum Information*, Phys. Rev. Lett. **109**, 050504 (2012).