Quantum error correction stores information in a subspace of a larger physical Hilbert, and is an efficient method of protecting quantum information from noise. Repeated measurements and error corrections keep the information from drifting too far out of the error correction subspace, also called the codespace. For robust quantum computation, we must be able to perform gates without leaving a protected codespace or amplifying existing errors. Fault-tolerance is straightforward for a limited set of gates, the so-called transversal gates of the code. For encoded single-qubit gates, transversal gates are accomplished by simply applying product unitaries. Unfortunately, severe constraints exist [1–4] that mean such direct approaches cannot provide gates sufficient for universal quantum computation. Rather we must rely on additional techniques to implement further fault-tolerant gates.

One route to universality is to prepare high-fidelity resource states, and then use state-injection to convert the resource state into a fault-tolerant gate [5–11]. Reduction of noise in these resource states, sometimes known as magic states [8], requires extensive distillation methods demanding that the vast majority of a quantum computer is a dedicated magic state factory [12, 13]. Due to the significant resource overhead, maximizing efficiency of these protocols is of paramount importance, and recently many improvements have been made [14–16]. One could try to circumvent this overhead by exploring one of many other ways to achieve universality [17–22]. However, all these proposals have to sacrifice some of the error correcting capabilities, and so are only viable when physical operations are much less noisy (such as was explicitly shown in Ref. [19]). Except for Shor’s method [17], these alternative routes require codes with a rare property, and such codes also play a fundamental role in most magic state distillation (herein MSD) protocols. Specifically, these codes have as a transversal gate the $\pi/8$ phase gate, which is special as it is outside the Clifford group yet still closely related to it.

In almost every route to fault-tolerance, these exotic codes emerge as pivotal components. Here we tackle the problem of designing, and improving, analogous codes in the qudit setting of using $d$-level elementary systems. In this setting, qudit error correction [23–25] has been long known, but only much later were analogs of $\pi/8$ phase gate characterized [26] and codes discovered with these as transversal gates [27]. Campbell, Anwar and Browne (herein CAB), analyzed several quantifiers of performance for magic state distillation using these codes [27, 28]. CAB found marked improvement over comparable qubit codes for modest size dimensions $d = 3, 5$, and such gains would carry over should these codes be employed in qudit analogs of the “magic-state free” fault-tolerance schemes [18–22]. However, in even larger dimensions ($d > 5$) performance again declined. This fall in performance is peculiar, especially in light of other results showing qudit toric codes have thresholds monotonically increasing with system dimension [29, 30]. Our protocols remedy this situation by providing a commensurate improvement to MSD, and together with qudit toric codes present an appealing and effective architecture. Toric codes provide fault-tolerant Clifford gates [31] and our protocols extend this to a universal set of fault-tolerant gates, with both components becoming more effective in higher dimensions.

We consider an extended class of quantum Reed-Muller codes, similar to those considered by CAB. Reed-Muller codes are constructed in terms of polynomial functions, and the work of CAB only considered linear functions. Here we see that the higher degree polynomials can be used to construct Reed-Muller codes with the desired transversality properties. Furthermore, the polynomial degree can grow with the qudit dimension $d$, and in turn the effectiveness of the code also grows. We make this statement more precise by considering the code parameters conventionally labelled $[[n, k, D]]_d$ where $n$ is the number of physical qudits, $k$ is the number of logical qudits encoded and $D$ is the code distance (measuring its error correction capabilities). The codes presented here encode a single qudit ($k = 1$) into $n = d−1$ qudits, and our best performing codes have $D = \lfloor (d+2)/3 \rfloor$, where $\lfloor \rfloor$ denotes the floor function. It is desirable that $D$ is larger and $n$ is smaller, but we find both numbers grow with system dimension. The overall effectiveness is measured by the codes “gamma value”, $\gamma = \log(n)/\log(D)$, which is smaller for more efficient codes. We find $\gamma$ can be decreased arbitrarily close to unity by increasing $d$, and discuss the operational meaning of $\gamma$ in the context of MSD. This puts qudit protocols far ahead of the first proposed qubit codes [8], and comparable with modern qubit block codes [14–16] without the disadvantages incurred by using block codes.

Aside from the practical merits, refinements to notation and proof techniques present a clearer picture of why these codes possess their strange properties. For reasons that become clear we limit ourselves to prime dimensions of 5 and above. We remark that extensions to prime power dimensions are plausible, and some techniques exist that offer hope in arbitrary dimensions [32]. Our results also contribute to our understanding of qudit magic states as a resource theory. Study of
the Pauli group, in particular Pauli-metric is not always invertible, which we need for the whole Pauli group. Multiplication in modular arithmetic is performed modulo $d$, so $\mathbb{F}_d$ forms a Galois field of order $d$. Next, let us review the structure of the Clifford group, denoted $C_d$, and a normal subgroup called the Pauli group, denoted $P_d$, both of which are fundamental to quantum coding theory. Those Clifford unitaries that are also diagonal (in the standard basis) have the form

$$Z_{\alpha, \beta}|x\rangle = \omega^{\alpha x + \beta x^2}|x\rangle,$$  

(1)

where $\alpha, \beta \in \mathbb{F}_d$ and throughout $\omega = \exp(i2\pi/d)$. Take note of the quadratic dependence of the phase. Whenever, $\beta = 0$ the exponent is linear in $x$ and the operator is also in the Pauli group, in particular Pauli-Z is $Z := Z_{1,0}$. Making use of the number operator $\hat{n} = \sum x|x\rangle\langle x|$, we could also write the above as $Z_{\alpha, \beta}|x\rangle = \omega^{\alpha \hat{n} + \beta \hat{n}^2}$. There are also Clifford unitaries that permute the computational basis states, such that

$$X_{\alpha, \beta}|x\rangle = |\alpha + \beta x\rangle,$$  

(2)

where again $\alpha, \beta \in \mathbb{F}_d$. Here $\beta = 1$ picks out elements of the Pauli group, in particular Pauli-X is $X := X_{1,1}$. The Pauli operators $X$ and $Z$, and tensor products thereof, generate the whole Pauli group. Multiplication in modular arithmetic is not always invertible, which we need for $X_{\alpha, \beta}$ to be unitary, but thankfully in prime dimensions we do have invertibility. These gates are not yet sufficient to generate the Clifford group and we must also include a Hadamard-like gate, $H$, that acts as $H|x\rangle = \sum y \omega^{xy}|y\rangle/\sqrt{d}$ and a 2-qudit control-phase gate of the form $C_Z = \omega^{\hat{\alpha} \hat{n}}$. Non-Clifford gates.-As remarked earlier, the special ingredient we need is a fault-tolerant implementation of a gate outside of the Clifford group. The qubit $\pi/8$ gate is non-Clifford and has other useful properties. We consider qudit analogs [26, 27]. Such an analog will be diagonal in the computational basis, non-Clifford, and will by conjugation map Pauli operators to Clifford operators. More formally, we require a diagonal unitary $M \notin C_d$ such that for all Pauli $P \in P_d$, we have $MPM^\dagger \in C_d$. Such gates are often said to belong in the third level of the Clifford hierarchy, and this property is useful for gate teleportation (a.k.a. state-injection) [42]. We show these properties for unitaries of the form, $M_{\mu} = \omega^{\mu \hat{n}^3}$ for $\mu \in \mathbb{F}_d \setminus \{0\}$, where $M$ stands for "magic". We could include a quadratic component to the exponent, but the unitary would be Clifford equivalent (for more insights on Clifford equivalence see Ref. [41] and App. (B)). An important point is that since $\omega^d = 1$, the exponent of $\omega$ can be calculated in modular arithmetic. Using $\omega^{\mu \hat{n}^3}X = X\omega^{\mu(\hat{n}+1)^3}$, we find:

$$M_\mu XM_\mu^\dagger = X\omega^{\mu((\hat{n}+1)^3-\hat{n}^3)} = X\omega^{\mu(3\hat{n}^2+3\hat{n}+1)},$$  

(3)

is Clifford and not Pauli. It follows that $M_{\mu}$ is in the third level of the Clifford hierarchy, and so analogous to a $\pi/8$ gate (more details in App. A).

The Reed-Muller codes.-Here we consider a simple subclass of shortened quantum Reed-Muller codes (herein QRM codes). The codes are defined in terms of polynomial functions from the non-zero elements of the field $(\mathbb{F}_d^* = \{1, \ldots, d-1\})$ to the whole field, so formally $F : \mathbb{F}_d^* \rightarrow \mathbb{F}_d$. It is called a degree-$r$ polynomial if

$$F(x) = f_0 + \sum_{m=1}^r f_m x^m,$$  

(4)

with $f_r \neq 0$ and where lower case is used throughout for the coefficients $f_k \in \mathbb{F}_d$. We also denote the degree of $F$ as $\deg(F) = r$. Fermat's little theorem (FLT) asserts that $x^n = x^m \pmod{d}$ if $n = m \pmod{d-1}$, and so all functions can be represented with polynomials of degree less than $d-1$. When higher degree polynomials appear, we equate them via FLT with the polynomial of lowest possible degree. We say the polynomial is unshifted if $f_0 = 0$, otherwise we say the function is shifted by $f_0 \neq 0$. We use these functions to describe quantum states, so that

$$|\psi_F\rangle = \bigotimes_{x=1}^{d-1} F(x).$$  

(5)

This notation can be unfamiliar at first so we provide some examples in Table. (I). More generally, Reed-Muller codes can be defined over a larger number of qudits by making use of polynomials over a larger domain (so multivariate polynomials, $F : \mathbb{F}_d^n \setminus \{0\} \rightarrow \mathbb{F}_d$), but such generality is not necessary for our purposes.

| $F(x)$ | $\deg(F)$ | $f_0$ | $|\psi_F\rangle$ |
|---|---|---|---|
| $x$ | 1 | 0 | $|1\rangle|2\rangle|3\rangle|4\rangle$ |
| $x^2$ | 2 | 0 | $|1\rangle|4\rangle|4\rangle|1\rangle$ |
| $2x+1$ | 1 | 1 | $|3\rangle|0\rangle|2\rangle|4\rangle$ |
| $x+x^2$ | 2 | 0 | $|2\rangle|1\rangle|2\rangle|0\rangle$ |

The codespace of such a QRM code is defined by its degree, $r$. We begin by defining the logical state $|0_L\rangle$ as

$$|0_L\rangle = d^{-r/2} \sum_{\deg(F) \leq r} |\psi_F\rangle,$$  

(6)

which is an equally weighted sum over all $|\psi_F\rangle$, where $F$ is an unshifted function of degree no greater than $r$. The other logical states are

$$|k_L\rangle = d^{-r/2} \sum_{\deg(F) \leq r} |\psi_F\rangle,$$  

(7)

where now the functions are instead shifted by $k$.

Alternatively, QRM codes can be described in the qudit stabilizer formalism [43, 44]. If a Pauli operator $s$ satisfies
sum over polynomials shifted by \( k \) for all \( k \), we say that \( s \) is a stabilizer of the code and write \( s \in S \). We proceed by defining Pauli operators

\[
X_F = \bigotimes_{x=1}^{d-1} X^{F(x)},
\]

and noting that \( X_F |\psi_G\rangle = |\psi_{F+G}\rangle \). One can then verify that the codespace is stabilized by \( X_F \) for all unshifted functions of degree no greater than \( r \). Similarly, other stabilizers of the form \( Z_F \) can be found, but where the degree is constrained to \( d - r - 2 \) (as we see later when discussing code distance). In terms of logical operators, we observe that since \( X^{\otimes(d-1)} |\psi_F\rangle = |\psi_{F+1}\rangle \), we have \( X^{\otimes(d-1)} |k_L\rangle = |(k+1)L\rangle \) and can identify \( \bar{X} = X^{\otimes(d-1)} \) where bars throughout denote logical operators.

Transversal gates.-The key feature of \( \mathcal{QRM} \) codes that makes them useful for MSD, and other approaches to fault-tolerant quantum computing, is that they possess transversal non-Clifford gates. That is, one can perform a logical non-Clifford gate within the codespace by applying a product unitary. We shall show that for \( \mathcal{QRM} \) codes of degree \( 3r < d - 1 \), a logical \( M \) gate can be implemented transversally by \( \overline{M}_\mu = M^{\otimes(d-1)} \). Notice how the proof here is again only meaningful for \( d > 3 \).

We begin by considering some polynomial \( F \), the corresponding state, \( |\psi_F\rangle \), and how the product unitary acts on this

\[
\overline{M}_\mu |\psi_F\rangle = \omega^{-\mu} \sum_{\bar{f} = 1}^{d-1} F(\bar{f})^3 |\psi_F\rangle = \omega^{-\mu S(H)} |\psi_F\rangle,
\]

where we introduce the shorthand \( S(H) := \sum_{x=1}^{d-1} H(x) \) and \( H(x) := F(x)^3 \), and next we must evaluate \( S(H) \). The following steps all rest on an remarkable algebraic feature (see App. C) of primes dimensions, namely that all functions satisfy \( S(H) = -h_0 \). Now we must find the explicit form for \( h_0 \) in terms of the \( f_m \). By expanding out \( F^3 \) and using FLT, we find \( h_0 \) is a sum over every \( f_{m_1} f_{m_2} f_{m_3} \) where \( m_1 + m_2 + m_3 = 0 \) (mod \( d-1 \)). This is hugely simplified if we restrict to \( F \) with degree less than \( (d-1)/3 \) as there is only one contribution such that \( h_0 = f_0^3 \). Under this assumption, Eq. (9) becomes simply \( \overline{M}_\mu |\psi_F\rangle = \omega^{-\mu f_0^3} |\psi_F\rangle \), and so the phase depends only on the shift of the function.

In Reed-Muller codes, the logical basis states \( |k_L\rangle \) are a sum over polynomials shifted by \( k \). Furthermore, the degree of these polynomials is no more than some degree \( r \). We have assumed \( 3r < d - 1 \) and so the above proof directly entails

\[
\overline{M}_\mu |k_L\rangle = \omega^{-\mu k^3} |k_L\rangle.
\]

This shows that the product unitary acts on the \( \mathcal{QRM} \) codewords as a logical \( M \) gate, and so transversality of a non-Clifford gate has been demonstrated.

Error correcting properties.- The previous work of CAB [27] used different proof techniques to show the transversality of non-Clifford gates for \( \mathcal{QRM} \) codes of only first degree, \( r = 1 \). Furthermore, those first degree codes could only detect a single error. However, increasing the degree of \( \mathcal{QRM} \) codes opens the possibility of detecting more errors with the dimensionality of the system. Indeed, we will show below that \( \mathcal{QRM} \) codes of degree \( r \) can detect up to \( r \) errors.

We first review some basic concepts. The weight of an operator \( P \), denoted \( \text{wt}(P) \), is the number of qudits it acts upon non-trivially. We are interested in the smallest weight Pauli operator whose effects cannot be detected by measuring stabilizers of the code and that also acts non-trivially on the code. Formally, \( P \) must commute with \( S \), but not be a member of \( S \), and the minimum weight of such \( P \) is called the code distance,

\[
D = \min\{\text{wt}(P)|[P,S] = 0, P \notin S, P \in \mathcal{P}\}.
\]

We begin by considering only phase errors, and again use polynomials to define multi-qudit operators, so \( Z_G = \otimes_x Z^G(x) \). The weight of \( Z_G \) can be expressed as \( \text{wt}(Z_G) = (d-1) - \chi(G) \), where \( \chi(G) \) is the number of non-zero arguments for which the function \( G \) evaluates to zero. That is, \( \chi(G) \) is the number of roots of the polynomial. Except for the trivial function \( (G(x) = 0) \), the roots are limited by the degree of the polynomial such that \( \chi(G) \leq \deg(G) \). Putting this together, we have the degree-weight relation \( \text{wt}(Z_G) \geq (d-1) - \deg(G) \). Next, we show that commutation of \( Z_G \) with the stabilizer puts an upper bound on the degree of \( G \).

Recall that the stabilizers of \( \mathcal{QRM} \) codes include operators \( X_F \) for all unshifted functions with \( \text{deg}(F) \leq r \). From \( ZX = \omega ZX \), we know that \( X_F \) and \( Z_G \) commute iff \( \sum_{x \in \mathbb{F}_p^d} F(x)G(x) = 0 \). Recall that such sums only vanish when the composite polynomial, here \( H'(x) := F(x)G(x) \), is unshifted. The shift of \( H' \) is \( h_0' \), which by expanding \( FG \) is a sum over every \( f_m g_n \), such that \( m + n = 0 \) (mod \( d-1 \)). Let us just consider monomials \( F(x) = x^q \), then we have the simplification \( h_0' = f_q g_{(d-1)-q} = g_{(d-1)-q} = -q \). By definition a degree \( r \) polynomial has \( g_r 
eq 0 \), and so \( h_0' \neq 0 \) whenever \( q = (d-1) - \deg(G) \). Since \( q = \deg(F) \leq r \), we conclude that provided \( (d-1) - r \leq \deg(G) \) we can always find an unshifted \( F \) (with \( \text{deg}(F) \leq r \)) such that \( FG \) is shifted. This entails that the corresponding \( Z_G \) fails to commute with at least one element of the stabilizer, namely \( X_F \). Conversely, for a \( Z_G \) error to commute with the stabilizer, it must have degree less than \( (d-1) - r \), and from the degree-weight relation this entails \( \text{wt}(Z_G) > r \). Being true for all \( Z_G \) entails all undetectable phase errors have weight greater than \( r \). As for \( X \) errors, a similar analysis shows even greater protection. We conclude that the code distance satisfies \( D \geq r + 1 \), with equality being easy to confirm.

MSD.- We have shown transversal non-Clifford gates for \( \mathcal{QRM} \) codes of up to degree, \( r = \lfloor (d-1)/3 \rfloor \), and so a distance, \( D = r + 1 = \lfloor (d + 2)/3 \rfloor \), that grows linearly with the system dimension. These results imply increasing code performance with increasing dimensionality. To make this concrete, let us consider the performance when using these codes for MSD. Each round of distillation consumes \((d-1)\) noisy copies of \( M_{\frac{d}{3}}[+] \) and, when successful, outputs a single purer magic state \( M_{\frac{d}{3}}[+] \). To briefly review, a simple distillation protocol will take noisy magic states \( \rho_M^{\otimes(d-1)} \), measure the stabilizers of a codespace and postselect on the “+1” outcomes. This results in a projection \( \Pi \) onto the codespace, and so we have \( \Pi \rho_M^{\otimes(d-1)} \Pi \).
Next, a decoding Clifford unitary is applied mapping $|k_L⟩ → |k⟩0 ⊗ (d−2)$, leaving the first qudit as a higher fidelity magic state. The role of transversality of $M_μ$ is the following, it entails $\Pi_ρ^{M_μ} \Pi = \Pi_ρ^{M_μ} \Pi (\Pi_ρ^{M_μ})^†$ where $ρ_X$ is a noisy state in the $X$-basis. This allows the stabilizer projection $Π$ to act on the noisy stabilizer state $ρ_X^{(d−1)}$ and directly detect errors therein. We expand on this last point. Writing $Π = Π_XΠ_Z$ where $Π_X$ and $Π_Z$ are projectors for the $X$ and $Z$ stabilizers of the code. We find $Π_Zρ_XΠ_Z$ lies in the codespace and is a logical $X$ state with some $Z$ noise. The subsequent $Π_X$ projection removes all detectable $Z$ errors, so that the state is output with an error rate $ε' = O(ε^2)$ where $D$ is the code distance, as analyzed in detail in Refs. [8, 27]. The above argument gives sufficient conditions for a code being useful for MSD. Codes without the desired transversality might, via a different mechanism, achieve MSD but the per- formance is no longer linked to code distance. For instance, MSD can be based on the 5-qubit code [8] or Steane code [9] and respectively reduces errors quadratically $ε' = O(ε^2)$ and linearly $ε' = O(ε)$, even though both codes have $D = 3$. Furthermore, prior work [8, 27] also shows that some “1” measurement outcomes can be accepted provided, as we have here, that $M_μ$ is in the third level of the Clifford hierarchy. In the analysis, we assume Clifford unitaries and Pauli measurements are implemented perfectly since an underlying fault-tolerance scheme can provide protection for these operations.

We now discuss the efficiency and noise threshold of MSD with the high degree $QRM_2$ codes. On average, the consumed number of noisy states can be shown [27] to scale as $C' log^2_2(ε_{final})$ where $γ = log_D(d−1)$. Therefore, the $γ$ value provides a good quantifier for the efficiency of protocols for MSD, with lower $γ$ values showing better performance. Using the optimal $D = [(d + 2)/3]$, we have

$$γ = \frac{\ln(d−1)}{\ln([(d + 2)/3])}. \quad (11)$$

As $d$ increases, we find $γ$ approaches 1 from above and we show some $γ$ values in Fig. (1a).

In prior analyses of qudit codes [27], only first degree $QRM_2$ codes were shown to have transversal non-Clifford gates, and such codes have constant distance $D = 2$. Furthermore, Ref. [27] observed improved $γ$ values up to dimension $d = 5$, followed by declining performance in larger dimensions. Whereas, here we find $γ$ values can continue to improve with dimensionality. In the qubit setting, the 15 qubit Reed-Muller codes achieves $γ = 2.46497$. However, recent improvements in qubit MSD have used block codes [14–16] that can also approach $γ = 1$ with increasing block size. A disadvantage in using block codes is that they need large complex circuits and many magic states must be simultaneously prepared to fully exploit the gains in efficiency. One comprehensive study of block code distillation within a surface code architecture showed the benefits of block codes are very slight [45]. By contrast, higher dimensional systems offer efficient protocols even for preparation of a single magic state, avoiding the complexity of block codes. It is natural to ask whether block codes exist for qudits, and whether they might be able to break below the $γ = 1$ barrier or obey a generalization of the $γ ≥ 1$ conjecture made by Bravyi and Haah [14].

Another important metric for MSD is the noise threshold below which the protocol successfully reduces noise. The threshold depends on the noise model, and here we consider depolarizing noise such that $M_μ|+⟩$ is mixed with the identity, and we quantify the noise by infidelity $ε = 1 − ⟨+|M_μρ_M|+⟩$. After a round of distillation, we again have a depolarized magic state of infidelity $ε' < ε$ provided $ε < ε_{dep}'$. For prime $d ≤ 17$, we found closed formulae for $ε'$ and solved for $ε_{dep}'$ as shown in Fig. (1), and observed monotonically increasing improvements with $d$.

**Conclusions.** In summary, we have shown that quantum Reed-Muller codes provide effective means of fault-tolerantly implementing gates that are essential to a variety of approaches to fault-tolerant quantum computing, with special attention paid to MSD. Unlike previous work, improvements in efficiency and thresholds continued into higher dimensions. We must remark that coherent control of high $d$ qudits is more challenging, and in physical systems one may see noise rise comparatively with the dimensionality. Such features depend subtly on the details of the underlying physics. Whilst many systems may not be well suited to qudit approaches, many atomic systems come equipped with large Hilbert spaces for which control of many levels need not be substantially more difficult than control of just 2 levels. For instance, experiments in trapped Cesium have performed gates between 16 levels at 99% fidelity [46].

**Acknowledgments.** We thank Mark Howard, Niel de Beau- drap, Hussain Anwar, Dan Browne and Tom Bullock for use- ful discussions and comments on the manuscript.
Actually the toric code possesses transversal gates that are a group.

This technicality can be overcome with an intermediate level of state distillation that provides the missing gates of the Clifford group.

In finite fields, multiplication is always invertible. Also, the zero element maps to zero, and so no other element can do so. Hence, typically $3x \neq 0$ when $x \neq 0$. This argument does not apply when $d = 3$ since multiplication by 3 is actually multiplication by 0, and so not invertible.

**Appendix A: Clifford hierarchy**

For completeness, we give extra details here on why $M_\mu$ (for nonzero $\mu$) is in the third level of the Clifford hierarchy. First we examine Eq. (3). Expanding the exponent, $t := (\hat{n} + 1)^3 - \hat{n}^3$, the cubic terms cancel and we are left with a purely quadratic expression, namely $t = \mu (3\hat{n}^2 + 3\hat{n} + 1)$. Therefore, the RHS is a Clifford that we denote $C = M_\mu X M_\mu^\dagger$. We further require that $M_\mu P M_\mu^\dagger$ is Clifford for all Pauli $P$. All Pauli operators can, up to a phase, be written $P = X^m Z^n$ and so conjugate to $M_\mu P M_\mu^\dagger = C^m Z^n$ where we have used commutation of $M_\mu$ with $Z$. Since the Clifford gates form a group $C^m Z^n$ is again Clifford.

Next we must check that $M_\mu$ is not simply a Clifford unitary. Clifford unitaries conjugate Pauli operators to Pauli operators. Therefore non-Cliffordness can be established by finding a single Pauli such that $M_\mu P M_\mu^\dagger$ is not a Pauli. We again consider $P = X$, and the RHS of Eq. (3). For $C$ to be non-Pauli, the exponent $t = \mu (3\hat{n}^2 + 3\hat{n} + 1)$ must be non-linear. This occurs provided the coefficient of $\hat{n}^2$ is nonvanishing. Therefore, we need $3\mu \neq 0$ modulo $d$. If $d = 3$, we have a problem as $3\mu \pmod{3}$ vanishes for all $\mu$. Whereas, for odd prime $d > 3$, multiplication by 3 is always non-trivial [47]. This is one of the fundamental reasons why it is simpler to construct codes with transversal non-Cliffords in odd $d > 3$.

**Appendix B: Clifford equivalence of $M_\mu$ gates**

Earlier we remarked that a $M_\mu$ gate will be Clifford equivalent to gates of the form $e^{\lambda \hat{n}^2 + \mu \hat{n}^3}$, simply by combining $M_\mu$ with a Clifford gate $Z_{\alpha, \beta}$. However, this prompts the question whether all $M_\mu$ are Clifford equivalent or whether...
the above $QR,M$ codes ever have genuinely distinct non-Clifford gates. This question was addressed in Ref. [41] in relation to mutually-unbiased bases (MUB) and symmetric, informally-complete (SIC) measurements, and here we present a brief review. In dimension $d = 2 \pmod{3}$, all $M_{\mu} \neq 0$ gates are Clifford equivalent. Whereas in dimensions where $d = 1 \pmod{3}$, we find 3 distinct equivalence classes.

We can move between different $M_{\mu}$ gates by conjugating the unitary with the permutation Clifford $X$ so that

$$M' = X_{\alpha,\beta} M_{\mu} X_{\alpha,\beta}^\dagger,$$

(B1)

where in the last line we collect all the quadratic terms into $g(n) = \mu \left( \alpha^3 + 3 \alpha^2 \beta n + 3 \alpha \beta^2 n^2 \right)$. The interesting part is the cubic term, which has gone from $\mu x^3$ to $\beta^3 \mu x^3$. Hence, $M_{\mu}$ and $M_{\mu'}$ are Clifford equivalent if there exists a $\beta$ such that $\mu' = \beta^3 \mu$. The structure of the equivalence class is determined by the set

$$R_d = \{ \beta^3 | \beta \in \mathbb{F}_d \setminus \{0\} \},$$

(B2)

which in field theory is known as the cubic residue of the field. It can be shown that the cubic residue forms a group under multiplication. We can now immediately leverage results in field theory.

In dimensions satisfying $d = 2 \pmod{3}$, the cubic residue includes all non-zero elements of the finite, so $R_d = \mathbb{F}_d \setminus \{0\}$. Therefore, we can always find a $\beta$ such that $\beta^3 = \mu / \mu'$ to ensure Clifford equivalence.

In dimensions satisfying $d = 1 \pmod{3}$, elementary field theory shows the cubic residue contains $(d-1)/3$ elements. The cubic residue is a normal subgroup of the group $\mathbb{F}_d \setminus \{0\}$ (a group under multiplication). Therefore, each element of $\mathbb{F}_d \setminus \{0\}$ belongs to one coset of the cubic residue, and each coset is equal in size. This gives three cosets each containing $(d-1)/3$ elements. Cosets are closed under multiplication by elements from the generating subgroup, the cubic residue, and hence each coset defines an equivalence class of $\mu$ values. For example, $d = 7$ satisfies $d = 1 \pmod{4}$ and the cubic residue is $R_7 = \{1,6\}$ with cosets $\{2,5\}$ and $\{3,4\}$, providing the three Clifford equivalence classes of $\mu$ values.

So far we have only considering changed $M_{\mu}$ by applying Clifford gates that permute and apply phases in the computational basis, having ignored the effect of the Hadamard gate. However, Hadamard gates simply change the basis, and so Hadamards effectively just interchange the role of subsequent $X_{\alpha,\beta}$ and $Z_{\alpha,\beta}$ Clifford unitaries.

**Appendix C: Evaluating summations**

We wish to find a general solution of

$$S(H) = \sum_{x \in \mathbb{F}_d} H(x),$$

(C1)

for all $H(x) = \sum_m h_m x^m$, and as usual working modulo $d$. We begin by breaking the sum into monomials,

$$S(H) = \sum_m h_m \left( \sum_{x \in \mathbb{F}_d} x^m \right) = \sum_m h_m S(x^m),$$

(C2)

and see the problem reduces to finding the monomial solutions $S(x^m)$.

First we consider sums for $m \neq 0 \pmod{d-1}$. Take a non-zero $y$, such that $y^m \neq 1$. The sum can be reordered to be over $(yx)^m$, so that

$$S(x^m) = \sum_{xy \in \mathbb{F}_d} (yx)^m,$$

(C3)

we reorder back

$$S(x^m) = y^m \sum_{x \in \mathbb{F}_d} x^m,$$

(C4)

and so $S(x^m) = y^m S(x^m)$. Since $y^m \neq 1$, we conclude $S(x^m) = 0$. Next we consider $m = 0 \pmod{d-1}$, for which $x^m = 1$ for all $x \in \mathbb{F}_d$ (recall the star denotes the absence of zero so we never encounter $0^0$). The sum has $(d-1)$ terms and so obviously $S(x^m) = d-1 = -1$.

We have shown that

$$S(x^m) = \sum_{x=1}^{d-1} x^m = \begin{cases} 0 & \text{when } m \neq 0 \pmod{d-1}, \\ -1 & \text{when } m = 0 \pmod{d-1}. \end{cases}$$

(C5)

The modular dependence on $m$ comes from FLT. We deduce that if $H$ is unshifted, and so contains no $x^0$ terms, then $S(H) = 0$, whereas in general $S(H) = -h_0$. 

6