Sublogarithmic Distillation in all Prime Dimensions using Punctured Reed-Muller Codes

Tanay Saha[1,2] and Shiroman Prakash[2]

[1] *Department of Mathematics, Simon Fraser University, Burnaby, B.C., Canada*
[2] *Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra, India*

Abstract

Magic state distillation is a leading but costly approach to fault-tolerant quantum computation, and it is important to explore all possible ways of minimizing its overhead cost. The number of ancillae required to produce a magic state within a target error rate ϵ is $O(\log^{\gamma}(\epsilon^{-1}))$ where γ is known as the yield parameter. Hastings and Haah derived a family of distillation protocols with sublogarithmic overhead (i.e., $\gamma < 1$) based on punctured Reed-Muller codes. Building on work by Campbell *et al.* and Krishna-Tillich, which suggests that qudits of dimension $p > 2$ can significantly reduce overhead, we generalize their construction to qudits of arbitrary prime dimension p. We find that, in an analytically tractable puncturing scheme, the number of qudits required to achieve sublogarithmic overhead decreases drastically as p increases, and the asymptotic yield parameter approaches $\frac{1}{\ln p}$ as $p \to \infty$. We also perform a small computational search for optimal puncture locations, which results in several interesting triorthogonal codes, including a $[[519, 106, 5]]_5$ code with $\gamma = 0.99$.

Contents

1 Introduction

One of the most promising approaches to quantum fault tolerance is magic state distillation – in which one first implements very high accuracy Clifford gates; these are used to distill high accuracy 'magic states' from a larger number of lower accuracy magic states [5, 27]. These magic states are then used to perform high fidelity non-Clifford gates. Magic state distillation is estimated to be a major source of overhead[1], and reducing this overhead is great theoretical and practical importance. While most work focuses on qubits, (e.g., [37, 38, 4, 9, 10, 23, 21, 20, 22, 33]), there has been some significant work in magic state distillation with qudits of odd prime dimension p [43, 1, 8, 6, 25, 12, 41, 29, 35, 34, 26, 36]. Some of these works, particularly, [8, 7, 29, 36], suggest substantial reductions in overhead may be possible by increasing p. Here, we build on these works, by constructing new families of triorthogonal codes that allow for magic state distillation with sublogarithmic overhead, for qudits of arbitrary prime dimension p.

Triorthogonal codes, first defined in [4], are a special class of $[[n, k, d]]_p$ codes that admit a transversal non-Clifford gate from the third level of the Clifford hierarchy, known as a T gate [24], and can be used to distill a corresponding magic state. Distilling via a triorthogonal code with parameters $[[n, k, d]]$, one can get k output magic states with error rate $\epsilon_{out} = O(A_d \epsilon_{in}^d)$ using n input states of error rate ϵ_{in}, where A_d is the number of logical-Z operators in the code of weight d. Concatenating z times, one obtains k magic states with error rate $\epsilon_{\text{out}} = O(A_d^{\frac{d^z-1}{d-1}} \epsilon_{in}^{d^z})$ from $\frac{n^z}{k^{z-1}}$ noisy magic states. Thus, the ratio of input to output magic states scales with the desired error rate, ϵ_{out} as $O\Big(\log^\gamma\left(\frac{1}{\epsilon_{out}}\right)\Big)$ where the yield parameter, $\gamma = \frac{\log(n/k)}{\log(d)}$, characterizes the overhead [5].

A Reed-Muller code was used in [5] to construct a $[[15, 1, 3]]_2$ code, now known to be triorthogonal, with a transversal non-Clifford gate that allowed for distillation with $\gamma = 2.47$. [4] first formalized the concept of a triorthogonal code for qubits, and constructed a family of $[[3k + 8, k, 2]]_2$ codes with $\gamma \to \log_2 3 \approx 1.6$. Around the same time, the $[[15, 1, 3]]_2$ code of [5] was generalized to qudits in [8], who considered Reed-Muller codes of linear degree with a single puncture, and, subsequently, [6], who used Reed-Solomon codes of maximal degree with a single puncture to construct $[[n, 1, d]]_p$ codes with $\gamma \to 1$, as $p \to \infty$.

Magic state distillation protocols with $\gamma < 1$ have sublogarithmic overhead. It was conjectured [4] that $\gamma \geq 1$ for all protocols. This was disproved by [20, 22], who showed that punctured binary Reed-Muller codes give rise to triorthogonal codes with $\gamma < 1$, with an explicit example with $\gamma = 0.68$ – the size of such codes is, however, extremely large. [29] showed that, for qudits, Reed-Solomon codes of maximal degree with multiple punctures give rise to triorthogonal codes with $\gamma \to 0$ as $p \to \infty$ with much smaller block sizes, although the smallest p for which $\gamma < 1$ is $p = 23$.

[1] We should remark that, over the past decade, there have been several notable advances that may substantially reduce this overhead cost in practice (e.g., [32, 11, 3, 16].)

More recently, families of binary triorthogonal codes with $\gamma \to 0$ have appeared [45, 17], perhaps making the question sublogrithmic distillation somewhat obsolete. Nevertheless, distillation protocols with sublogarithmic distillation are exceedingly rare, and seem to require inordinantly large block sizes, or qudit dimensionalities. It is therefore appears worthwhile to explore all possible constructions of such codes – particularly, to help more precisely quantify the possible advantages of working with qudits over qubits. Moreover, there exists a very simple block construction of a family of small qutrit triorthogonal codes [36], for which $\gamma \to 1$. It appears to be a meaningful question to search for triorthogonal codes that outperform this simple family.

Here, we present a natural extension of the work of [8, 6, 29] and generalize the constructions of [22, 20] to qudits, to study qudit triorthogonal codes for all prime dimensions using punctured Reed-Muller codes of maximal degree. We construct a puncturing scheme in which the parameters of the resulting triorthogonal code are analytically computable, and show that these codes distill magic states with $\gamma < 1$ for qudits of any prime dimension. The size of the code, n, required to achieve $\gamma < 1$ decreases drastically with p, as p increases from 2 to 23. In the asymptotic limit of large n, the yield parameter of our codes vanishes as $\frac{1}{\ln p}$ as $p \to \infty$. We also perform a randomized computer search over puncture locations to construct qutrit and ququint punctured Reed-Muller codes with high yields – this revealed a $[[519, 106, 5]]_5$ code with $\gamma = 0.99$. To our knowledge, all explicit constructions of binary triorthogonal codes giving rise to sublogarithmic distillation require block sizes several orders of magnitude greater.

Our paper is organized as follows. In section 2, we review the construction of qudit triorthogonal codes from triorthogonal spaces. In section 3 we review Reed-Muller codes and discuss how they can be used to construct triorthogonal spaces, and then punctured to construct triorthogonal codes. In section 4, we present a puncturing scheme generalizing that of [22] that allows us to construct a family of $[[n, k, d]]_p$ triorthogonal codes, that give rise to distillation routines with sub-logarithmic yield when the codes are sufficiently large. An appendix contains more details about the asymptotic performance of this code. In section 5, we summarize results from a computational search for other puncturing schemes that give rise to qutrit and ququint triorthogonal codes with better parameters. In section 6, we conclude with some discussion regarding the possible advantages of fault-tolerant quantum computing with qudits rather than qubits.

2 Triorthogonal codes

First defined in [4] and generalized to qudits with prime dimension in [29], triorthogonal codes are CSS quantum error-correcting codes transversal in the T-gate, a diagonal gate from the third-level of the Clifford hierarchy, defined for qudits of arbitrary prime dimension in [24, 8]. Triorthogonal codes are constructed from (classical) triorthogonal spaces, as we now review.

Let $p \in \mathbb{N}$ be prime and let $\mathbb{F}_p$ denote the finite field with p elements. Define the $*$-product of two vectors $v = (v_1, v_2, \dots) \in \mathbb{F}_p^n$ and $w = (w_1, w_2, \dots) \in \mathbb{F}_p^n$ to be the vector $v * w = (v_1 w_1, v_2 w_2, \dots, v_n w_n)$.

Definition 1 (Classical triorthogonal space)**.** *A linear space $\mathcal{C}$ over the field $\mathbb{F}_p$ is classical tri-orthogonal if:*

1. $\forall x, y, z \in \mathcal{C}$, $\sum_i (x * y * z)_i = 0 \ (mod\ p)$
2. $\forall x, y \in \mathcal{C}$, $\sum_i (x * y)_i = 0 \ (mod\ p)$.

Note that, from its definition, any triorthogonal space is also self-orthogonal.

Let G be the generator matrix of a triorthogonal space. Let G' be the generator matrix of a triorthogonal space punctured in k locations, and let G_0 be the generator matrix of that space shortened in the same k locations,

$$G_{n_0 \times k_0} \xrightarrow[k \text{ columns}]{\text{Puncture @}} G' = \left(\begin{array}{ccc} \dots & r_1 & \dots \\ \dots & \vdots & \dots \\ \dots & r_k & \dots \end{array}\right) \Big\} G_1 \quad \left(\begin{array}{ccc} \dots & r_{k+1} & \dots \\ \dots & \vdots & \dots \end{array}\right) \Big\} G_0 \tag{1}$$

Let $\mathcal{G}'$ and $\mathcal{G}_0$ be the linear classical codes generated by these generator matrices. Clearly, $\mathcal{G}_0 \subseteq \mathcal{G}'$ and G' can be written as, $G' = \left(\frac{G_1}{G_0}\right)$. It is shown in [4] (and [29] for qudits) that the $CSS(G_0 \to X, G'^{\perp} \to Z)$ is a CSS code with the desired transversal T-gate, known as a triorthogonal code. Assuming that the submatrix of G that survives after puncturing has full rank, the length of the CSS code is $n_0 - k$ and the dimension is k. For non-degenerate[2] G_0 and G', the distance of this code is $\min(d(G_0^{\perp}), d(G')) = d(G_0^{\perp})$. (This is because, by triorthogonality, $G_0 \subseteq G'^{\perp}$, and therefore $G' \subseteq G_0^{\perp}$.)

3 Reed-Muller codes

Reed-Muller codes are constructed from the linear vector space of polynomials of total degree r in m variables over $\mathbb{F}_p$. They provide an elegant construction of triorthogonal codes, as observed in [5, 8, 20, 22], as we now review.

[2] If the code is degenerate, the distance of the code is the minimum Hamming weight of a vector in $\mathcal{G}_0/\mathcal{G}'^{\perp}$, which is greater than or equal to $d(\mathcal{G}_0^{\perp})$. The codes studied in this paper are always non-degenerate.

3.1 Triorthgonal spaces from Reed-Muller codes

Let $RM_p(r,m)$ denote the classical Reed-Muller code over $\mathbb{F}_p$. A codeword of $RM_p(r,m)$ is a complete list of function values $\mathrm{ev}(f) = (f(\vec{v}) : \vec{v} \in \mathbb{F}_p^m) \in \mathbb{F}_p^{p^m}$ where f is a polynomial of total degree at most r in m variables $(x_1, \ldots x_m)$ such that $x_i^p = x_i$. In other words, $RM_p(r,m) = \{f : \deg(f) \leq r, f \in \mathbb{F}_p[x_1, x_2, \ldots, x_m]/\langle \ldots, x_i^p - x_i, \ldots \rangle\}$. We do not distinguish between a codeword and the polynomial that gives rise to that codeword in $RM_p(r,m)$. Note, also, that $RM_p(r,m)$ is invariant under permutations of coordinates that correspond to affine transformations of $\mathbb{F}^m$, because the transformation, $\vec{r} \to M\vec{r} + \vec{r}_0$ does not change the degree of $f(\vec{r})$.

The degree r of a Reed-Muller code $RM_p(r,m)$ must satisfy

$$r \leq m(p-1), \tag{2}$$

because $x^p = x \mod p$.

It is a well-known fact that the dual of a Reed-Muller code is another Reed-Muller code.

Lemma 1. *$RM_p(r,m)^\perp = RM_p(\tilde{r}, m)$ where $\tilde{r} = m(p-1) - r - 1$.*

This follows from the fact that

$$\sum_{x,y \in \mathbb{F}_p} x^\alpha y^\beta \neq 0 \tag{3}$$

iff $\alpha = 0 \mod (p-1)$ and $\beta = 0 \mod (p-1)$.

As an immediate corollary of the above lemma, we have the following.

Theorem 1. *$RM_p(r,m)$ is a classical triorthogonal space if $3r < m(p-1)$.*

Proof. Let $r_{\max} = \lfloor \frac{m(p-1)-1}{3} \rfloor$. In finite fields, $\sum_{x \in \mathbb{F}_p} x^r = 0 \; \forall r < p-1$. So, $|ev(f_1) * ev(f_2) * ev(f_3)|_M = |ev(f_1 f_2 f_3)|_M = 0$ if $r < r_{\max}$, using the summation relation on all m variables. □

The distance of a classical Reed-Muller code $RM_p(r,m)$ is determined by polynomial in m variables of degree r with the most zeros. This is given by the Schwartz-Zippel lemma, [13, 40, 47], which we reproduce here,

Theorem 2. *The Schwartz-Zippel Lemma states that, given a multivariate polynomial $f(\vec{x})$ of degree r in m variables over $\mathbb{F}_p$, the total number of nonzero entries satisfies*

$$|\{\vec{x} | f(\vec{x}) \neq 0\}| \geq p^{m - \lfloor \frac{r}{p-1} \rfloor} \left(1 - \frac{r \bmod (p-1)}{p}\right). \tag{4}$$

Here, $r \bmod (p-1)$ denotes the the remainder when r is divided by $p-1$. For $p = 2$, $r \bmod (2-1) = 0$. This bound is tight, so the distance of a classical Reed-Muller code is,

$$d_{RM}(r,m) = p^{m - \lfloor \frac{r}{p-1} \rfloor} \left(1 - \frac{r \bmod (p-1)}{p}\right). \tag{5}$$

3.2 Puncturing Reed-Muller codes

We can form a triorthogonal code from this Reed-Muller space by puncturing it in a k different locations, which we denote as $S = \{\vec{x}_1, \ldots \vec{x}_k\}$.

Theorem 3. *The distance of the quantum triorthogonal code constructed from puncturing a $RM_p(r, m)$ triorthogonal space is $d(PRM_p(\tilde{r}, m; S))$ where $\tilde{r} = m(p-1) - r - 1$.*

Proof. $\mathcal{G}_0$ will then be a shortened Reed-Muller code, denoted as $SRM_p(r, m; S)$, while $\mathcal{G}'$ will be a punctured Reed-Muller code, denoted as $PRM_p(r, m; S)$. Recall that the dual of a shortened code is a punctured code, so $\mathcal{G}_0^\perp = SRM_p(r, m; S)^\perp = PRM_p(\tilde{r}, m; S)$, and $d(PRM_p(\tilde{r}, m; S))$ assuming non-degeneracy.

To see that degeneracy does not further increase the distance of the code, note that $\mathcal{G}'^\perp = PRM_p(r, m; S)^\perp = SRM_p(\tilde{r}, m; S)$. There always exists at least one codeword of minimum weight in $PRM_p(\tilde{r}, m; S)$ that is not be present in $SRM_p(\tilde{r}, m; S)$. A codeword of minimum weight d in $PRM_p(\tilde{r}, m; S)$ corresponds to a polynomial f in $RM((\tilde{r}, m)$. If f vanishes on all puncture locations, one can act on it with an affine transformation to obtain another polynomial f' that is non-zero on one of the puncture locations, which gives rise to codeword of weight $d' \leq d$ in $PRM_p(\tilde{r}, m; S)$. Either f or f' is nonzero on one of the puncture locations, and is therefore absent from $SRM_p(\tilde{r}, m; S)$. □

Campbell *et al.* [8] punctured $RM_p(1, m)$ once to construct triorthogonal codes for qudits. However, it is always advantageous to define a triorthogonal space using $RM_p(r, m)$, with r taken to be as large as possible, subject to $3r < p(m-1)$. To see this, note that $\tilde{r}(r)$ from Theorem 3 is a monotonically decreasing function of r. For $\tilde{r}_1 < \tilde{r}_2$, $PRM_p(\tilde{r}_1, m) \subset PRM_p(\tilde{r}_2, m)$, so $d(PRM_p(\tilde{r}_1, m)) \geq PRM_p(\tilde{r}_2, m)$.

Reed-Muller codes of maximal degree, $r = r_{\max} = \lfloor \frac{p(m-1)-1}{3} \rfloor$ therefore provide a natural construction of triorthogonal spaces. However, one is then faced with the question: how does one choose the set of puncture locations? For fixed k, one may wish to choose k puncture locations in order to maximize the distance d of the resulting triorthogonal code. Moreover, one also wishes to vary the number of puncture locations k in order to optimize γ. These appear to be non-trivial combinatorial questions.

3.3 Codes with one or two punctures

Let us first discuss the case of Reed-Muller codes with one or two punctures. Triorthogonal codes arising from Reed-Muller codes with one or two punctures do not give optimal yields. However, for a given m, they have the largest possible distance. The codes we construct in this subsection always outperform the punctured $RM_p(1, m)$ codes of [8], but need not outperform the $RM_p(r, 1)$ codes (i.e., Reed-Solomon codes) of [7].

We first have the following lemma.

Lemma 2. *The classical distance of a punctured Reed-Muller code $RM_p(r,m)$ with one or two punctures is independent of the locations of the punctures. Assume $d_{RM}(r,m) \geq 2$. For one puncture, the distance is $d_{RM}(r,m)-1$, and for two punctures the distance is $d_{RM}(r,m)-2$.*

Proof. $RM_p(r,m)$ is invariant under permutations of coordinates that correspond to affine transformations of $\mathbb{F}_p^m$. We can therefore, without loss of generality choose the first puncture to be at the point $(0,\ldots,0) \in \mathbb{F}_p^m$, and the second puncture to be at $(1,\ldots,0) \in \mathbb{F}_p^m$.

Consider any minimum weight codeword in the unpunctured Reed-Muller code, which corresponds to a polynomial f which is non-zero on at least 2 points in $\mathbb{F}_p^m$, by assumption. One can act on this polynomial with an affine transformation on $\mathbb{F}_p^m$ to obtain a minimal weight polynomial which is non-zero on both puncture locations. The distance of the code is therefore reduced by 1 for each of the first two punctures. □

Note that, for $p > 2$, there are several possibly-inequivalent choices for the location of the third and subsequent punctures, depending on whether the point $\vec{x}$ at which the third puncture is located is taken to be collinear with the first two punctures in $\mathbb{F}_p^m$. It is, therefore, also possible that, if the location is chosen carefully, the third puncture does not further reduce the distance of the resulting triorthogonal code.

The parameters of a triorthogonal code arising from puncturing $RM_p(r_{\max},m)$ once or twice are $[[p^m-k,k,d_{RM}(\tilde{r}(r_{\max},m)]-k]_p$, for $k \leq 2$.

Parameters of the smallest ternary Reed-Muller codes with a single puncture are $[[8,1,2]]_3$, $[[26,1,2]]_3$, $[[80,1,5]]_3$, $[[242,1,8]]_3$, $[[728,1,8]]_3$, $[[2186,1,17]]_3$, and $[[6560,1,26]]_3$. The first two codes in the list, which have distance 2, were constructed in [8], but the remaining are new. In the limit $m \to \infty$, the yield parameter of such codes goes to $\gamma \to 3$; but, the yield parameter for these codes may be irrelevant in practice, as one may only need to do a single round of distillation with codes of reasonably large distance (e.g., [15]). For $p = 5$, small codes include those with parameters, $[[4,1,2]]_5$, $[[24,1,3]]_5$, $[[124,1,4]]_5$, and $[[624,1,14]]_5$. For $p = 7$, we have $[[6,1,2]]_7$, $[[48,1,4]]_7$, $[[342,1,6]]_7$ and for $p = 11$, we have codes $[[10,1,4]]_{11}$, $[[120,1,7]]_{11}$. The first of these are the Reed-Solomon codes first proposed in [7].

As an example, consider the $[[80,1,5]]_3$. The noise reduction is plotted in Figure 1. Its threshold of the code to depolarizing noise, defined via $\rho_M(\delta) = (1-\delta)\,|M\rangle\langle M| + \delta\frac{1}{p}$, is $\delta* = 0.16$. In general, for qudits of dimension p there are $p-1$ distinct noise channels [8] – we also checked that it distills all types of noise, not just depolarizing noise.

For more than two puncture locations, it appears that one must carry out a randomized or exhaustive search over puncture locations to obtain triorthogonal codes of high distance – we carry out a modest search in section 5 similar to that provided for binary codes in [20]. Such searches are naturally limited to small codes over small primes p. In the next section, we present a puncturing scheme generalizing that used in [22] to all prime dimensions. This scheme, while suboptimal, allows one to compute the distance analytically, for arbitrarily

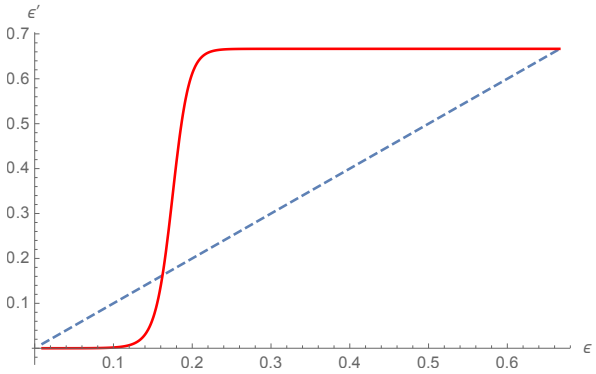

Figure 1: Distillation performance of the $[[80, 1, 5]]_3$ code, $\epsilon'(\epsilon)$ for depolarizing noise.

large codes.

4 Sublogarithmic distillation

In this section, we present an explicit scheme for puncturing triorthogonal spaces constructed from $RM_p(r, m)$ codes that give rise to triorthogonal codes that distill with sublogarithmic overhead for sufficiently large m for all p.

4.1 Puncturing schemes based on weight functions

Hastings and Haah [22] devised a puncturing scheme for the binary Reed-Muller code $RM_2(r, m)$ which, although not optimal, allows one to derive explicit combinatorial expressions for the distance of the punctured Reed-Muller code for arbitrary r and m. The key idea is puncturing in locations chosen according to the weight of the coordinate $\vec{x} \in \mathbb{Z}_2^m$ viewed as a binary codeword. One of the main results of the paper is that it is possible to generalize this idea to $RM_p(r, m)$. Unlike the binary case, however, there are many choices of the weight function possible.

Let $\vec{v} \in \mathbb{F}_p^m$ be a coordinate for $RM_p(r, m)$. A general weight function takes the form

$$|\vec{v}|_W = \sum_{i=1}^{n} W(v_i), \tag{6}$$

where $W : \mathbb{F}_p \to \mathbb{Z}$. The simplest example is the Hamming weight, $|\vec{v}|_H$, defined via

$$W_H(\alpha) = \begin{cases} 0 & \alpha = 0 \\ 1 & \alpha \neq 0 \end{cases}. \tag{7}$$

Another weight function that has appeared in the literature is, the Lee weight, $|\vec{v}|_L$ defined

via,

$$W_L(\alpha) = \begin{cases} \alpha & \alpha \leq (p-1)/2 \\ p-\alpha & \alpha > (p-1)/2 \end{cases}. \tag{8}$$

The Lee weight of a vector $\vec{v}$, is essentially the distance to the origin on a periodic $p \times p \times \ldots \times p$ lattice.

We introduce a new distance function, which we find is more convenient for our purposes than the previous two weight functions, denoted as $|\vec{v}|_M$, which we call the *Manhattan weight*, defined via,

$$W_M(\alpha) = \alpha. \tag{9}$$

$|\vec{v}|_M$ is the distance to the origin on a $p \times p \times \ldots \times p$ lattice that is not periodic using the Manhattan metric. In both the Lee and Manhattan weight functions, we implicitly convert $\alpha \in \mathbb{F}_p$ to $\alpha \in \{0, 1, 2, \ldots, p-1\} \subset \mathbb{Z}$.

Define a "generalized binomial coefficient", $C(m, k; W, p)$, via the generating function

$$\left(\sum_{j=0}^{p-1} x^{W(j)} \right)^m = \sum_{k=0}^{\infty} C(m, k; W, p) x^k. \tag{10}$$

$C(m, k; W, p)$ counts how many words there are of weight $\tilde{W}(\vec{v}) = k$ in the vector space $\mathbb{F}_p^m$. For Hamming weight, this takes the form

$$(1 + (p-1)x)^m = \sum_{j=0}^{m} C(m, j; W_H, p) x^j, \tag{11}$$

and $C(m, j; W_H, p) = (p-1)^j \binom{m}{j}$. For the Manhattan weight,

$$\left(1 + x + x^2 + \ldots + x^{p-1}\right)^m = \sum_{j=0}^{m(p-1)} C(m, j; W_M, p) x^j. \tag{12}$$

The puncturing scheme of [22], in this more general setting, is to puncture all coordinates with weight less than or equal to w, for some weight function. We denote the number of such coordinates by

$$C_{\leq}(m, k; W, p) = \sum_{j=0}^{k} C(m, j; W, p) \tag{13}$$

and also for convenience, define,

$$C_{>}(m, k; W, p) = \sum_{j=k+1}^{\infty} C(m, j; W, p) = p^m - C_{<}(m, k; W, p). \tag{14}$$

For any weight function, this gives rise to a triorthogonal code with parameters[3]

$$[[C_{>}(m, w; W, p), C_{\leq}(m, w; W, p), d(m, r, w; W, p)]]_p, \tag{15}$$

[3]This formula, (15), assumes that the punctured generator matrix for the classical triorthogonal space continues to be of full-rank, which is necessarily true if the number of puncture locations is less than the minimum weight of a codeword in the unpunctured triorthogonal space, i.e., if $C_{\leq}(m, w; W, p) < d_{RM}(r, m)$.

where $d(m, r, w; W, p)$ is the distance of a punctured Reed-Muller code, and depends non-trivially on both the choice of W, and the degree r used to define the Reed-Muller code.

None of these puncturing schemes is optimal. Note, however, that $|\vec{v}|_M \geq |\vec{v}|_L \geq |\vec{v}|_H$, and therefore $C(m, k; W_M, p) \leq C(m, k; W_L, p) \leq C(m, k; W_L, p)$. Therefore, of the weight functions listed above, puncturing according to Manhattan weight will give us the most freedom in choosing the total number of puncture locations – which can then be tuned to optimize γ. It turns out that $d(m, r, w; W_M, p)$ can be calculated in this case following the same strategy as [22]. We therefore focus exclusively on this case.

4.2 Combinatorial tools: p-nomial coefficients

Let us introduce some combinatorial tools to aid in counting sets of specified Manhattan weight.

We introduce the notation $\binom{m}{s}_p = C(m, s; W_M, p)$, i.e.,

$$(1 + x + x^2 + \cdots + x^{p-1})^m = \sum_{s=0}^{m(p-1)} \binom{m}{s}_p x^s. \tag{16}$$

We refer to $\binom{m}{j}_p$ as **p-nomial coefficents**.[4] Notice that when $p = 2$, the $p-$nomial coefficients are the familiar binomial coefficients.

Many of the identities satisfied by binomial coefficients generalize immediately to $p-$nomial coefficients. We list some of these below.

Consider,

$$(1 + x + \cdots + x^{p-1})(1 + x + \cdots + x^{p-1})^m = (1 + x + \cdots + x^{p-1}) \sum_{s=0}^{m(p-1)} \binom{m}{s}_p x^s \tag{17}$$

$$(1 + x + \cdots + x^{p-1})^{m+1} = \sum_{s=0}^{(m+1)(p-1)} \binom{m+1}{s}_p x^s. \tag{18}$$

Comparing the coefficients, we get the following lemma.

Lemma 3. *Generalized Pascal's Rule:*

$$\binom{m+1}{s}_p = \sum_{i=0}^{p-1} \binom{m}{s-i}_p. \tag{19}$$

Using the multinomial expansion, the $p-$nomial coeffients can be expressed as a sum of multinomial coefficients:

[4] See [44] and references therein for discussion of these coefficients for the case $p = 3$, which some authors call trinomial coefficients. p-nomial coefficients for general p were discussed by Euler in [14].

Lemma 4. *The coefficient of x^s in $(1+x+x^2+\cdots+x^{p-1})^m$ is*

$$\binom{m}{s}_p = \sum_{k_1+2k_2+\cdots+(p-1)k_{p-1}=s; k_i\geq 0} \frac{m!}{k_1!\,k_2!\ldots k_{p-1}!\,(m-k_1-k_2-\cdots-k_{p-1})!}. \tag{20}$$

In what follows, it is also convenient to introduce the following short-hand notation, following [22],

Definition 2. *We define*

$$\binom{m}{>s}_p = \sum_{i>s}\binom{m}{i}_p \text{ and } \binom{m}{\leq s}_p = \sum_{i\leq s}\binom{m}{i}_p. \tag{21}$$

Note that $\binom{m}{>s}_p \geq 0$ $\binom{m}{\leq s}_p \geq 0$ $\binom{m}{s}_p \geq 0$, and are only defined for positive m, s.

It is straightforward to derive the following lemma.

Lemma 5.

$$\binom{m}{>k}_p + \binom{m}{\leq k}_p = p^m. \tag{22}$$

In Appendix A, we discuss the asymptotic form of the $p-$nomial coefficients, $\lim_{m\to\infty}\frac{1}{m}\log_p\binom{m}{m\theta}_p = H_p(\theta)$.

4.3 Distance of the punctured Reed-Muller code

We now compute the minimum distance of the classical Reed-Muller code, punctured according to the Manhattan weight scheme.

Reed-Muller codes are defined as the evaluations of polynomials in $\mathbb{F}_p[x_1,x_2,\ldots,x_m]/\langle x_i^p - x_i\rangle$ over the vectors $\vec{x}\in\mathbb{F}_p^m$. Puncturing this code is equivalent to removing a subset $S\subset\mathbb{F}_p^m$ of points from $\mathbb{F}_p^m$. For our construction, we choose this subset to be $S_p(w) = \{v\in\mathbb{F}_p^m : |v|_M \leq w\}$, and we call the punctured code $PRM_p(r,m,w)$.

Each function $f\in\mathbb{F}_p[x_1,x_2,\ldots,x_m]/\langle x_i^p - x_i\rangle$ with total degree less than r corresponds to a codeword in $PRM_p(r,m,w)$. Let $|f|_{>w}$ denote the support of f in $\mathbb{F}_p^m \setminus S_p(w)$. $|f|_{>w}$ is the Hamming weight of the codeword corresponding to f in $PRM_p(r,m,w)$. The function with smallest non-zero $|f|_{>w}$ determines the minimum distance of the code.

Theorem 4. *The distance of the classical punctured Reed-Muller code, $PRM_p(r,m,w)$, is*

$$\begin{aligned}\Delta_p(m,r,w) &= \sum_{j=0}^{p-\beta-1}\binom{m-\alpha-1}{>w-j}_p \\ &= \begin{cases}\binom{m-\alpha}{>w}_p & \beta=0 \\ \binom{m-\alpha}{>w}_p - \sum_{j=p-\beta}^{p-1}\binom{m-\alpha-1}{>w-j}_p & \beta\neq 0,\ \alpha\leq m-1.\end{cases}\end{aligned} \tag{23}$$

where $r=\alpha(p-1)+\beta\leq m(p-1)$ with $\beta\in\{0,1,\ldots,p-2\}$, assuming that $\Delta_p(m,r,w)>0$.

Proof. We will show that any nonzero function of total degree r in $\mathbb{F}_p[x_1,\ldots,x_m]/\langle x_i^p - x_i\rangle$ has support in $\mathbb{F}_p^m \setminus S_p(w)$ greater than or equal to $\Delta_p(m,r,w)$, and that there is a function that saturates this inequality. If $\Delta_p(m,r,w) > 0$, then $\Delta_p(m,r,w)$ is the minimum distance of the punctured Reed Muller code, as per the discussion above.

Consider the polynomial

$$m_0(\vec{x}) = \prod_{i=1}^{\alpha}(1-x_i^{p-1}) \cdot \prod_{j=0}^{\beta-1}(p-1-i-x_{\alpha+1}) \in RM_p(r,m). \tag{24}$$

This polynomial has degree r and is thus a codeword in $PRM_p(r,m,w)$.

Let us show that $|m_0(\vec{x})|_{>w} = \Delta_p(m,r,w)$, which is the number of points in $\mathbb{F}_p^m \setminus S_p(w)$ where $m_0(\vec{x}) \neq 0$. From the form of $m_0(\vec{x})$, we see that any such point must have $x_1 = x_2 = \cdots = x_\alpha = 0$ and $x_{\alpha+1} \in \{0,1,\ldots,p-\beta-1\}$. Thus, the weight of this codeword is the number of solutions to $x_{\alpha+1} + x_{\alpha+2} + \cdots + x_m > w$ where $x_i \in \{0,1,\ldots,p-1\}$ for $i > \alpha+1$. This is given by $\sum_{j=0}^{p-\beta-1} \binom{m-\alpha-1}{>w-j}_p = \Delta_p(m,r,w)$.

We will now prove that no function f in $PRM_p(r,m,w)$ has $|f|_{>w} < \Delta_p(m,r,w)$. We will proceed by induction on the number of variables m, recalling that the degree must satisfy, $r = \alpha(p-1)+\beta \leq m(p-1)$. The base case of $m = 1$ (for which $r \leq p-1$) is trivial to show. Let $f^{(m)}$ be any polynomial in m variables such that $\deg(f) \leq \min(r, m(p-1))$. We assume, for the inductive argument, that $|f^{(m)}(x_1,x_2,\ldots,x_m)|_{>w} \geq \Delta_p(m,r,w)$. We now show that this assumption implies that any polynomial $f^{(m+1)}$ in $m+1$ variables, with $\deg(f^{(m+1)}) \leq \min(r,(m+1)(p-1))$, then $|f^{(m+1)}|_{>w} \geq \Delta_p(m+1,r,w)$.

Note that, if $r = (m+1)(p-1)$, the result is trivial, since $\Delta_p(m+1,(m+1)(p-1),w) = 0$. So we assume $\deg(f^{(m+1)}) = r < (m+1)(p-1)$ in what follows. (This does not, however, imply that $r < m(p-1)$.) Any $f^{(m+1)}(x_1,x_2,\ldots,x_{m+1})$ can be written as

$$f^{(m+1)} = f_0^{(m)} + f_1^{(m)}x_{m+1} + f_2^{(m)}x_{m+1}^2 + \cdots + f_{p-1}^{(m)}x_{m+1}^{p-1} \tag{25}$$

where $\deg(f_i^{(m)}) \leq \min(r-i, m(p-1))$. Thus, summing over all values of x_{m+1}, we have,

$$\begin{aligned} |f^{(m+1)}|_{>w} = &\left|f_0^{(m)}\right|_{>w} + \left|f_0^{(m)} + f_1^{(m)} + \cdots + f_{p-1}^{(m)}\right|_{>w-1} \\ &+ \left|f_0^{(m)} + 2f_1^{(m)} + \cdots + 2^{p-1}f_{p-1}^{(m)}\right|_{>w-2} + \ldots \\ &+ \left|f_0^{(m)} - f_1^{(m)} + f_2^{(m)} - \cdots + f_{p-1}^{(m)}\right|_{>w-(p-1)}. \end{aligned} \tag{26}$$

We denote each of the terms in the above equation as,

$$f^{(m+1)}|_{x_{m+1}\to z} = f_0^{(m)} + f_1^{(m)}z + f_2^{(m)}z^2 + \ldots f_{p-1}^{(m)}z^{p-1}. \tag{27}$$

(So, e.g., $f^{(m+1)}|_{x_{m+1}\to 0} = f_0^{(m)}$, and $f^{(m+1)}|_{x_{m+1}\to 2} = f_0^{(m)} + 2f_1^{(m)} + \cdots + 2^{p-1}f_{p-1}^{(m)}$.)

Let us first consider

- Case 1: $\forall z \in \mathbb{F}_p$, $f^{(m+1)}|_{x_{m+1}\to z}$ does not identically vanish.

  Each of the terms $f^{(m+1)}|_{x_{m+1}\to z}$ in eq. (26) has m variables and degree $r^* \leq \min(r, m(p-1))$.

  Then, by the induction hypothesis,

$$|f_0|_{>w} \geq \Delta_p(m, r^*, w) \geq \Delta_p(m, r, w) \tag{28}$$
$$|f_0 + f_1 + \cdots + f_{p-1}|_{>w-1} \geq \Delta_p(m, r^*, w-1) \geq \Delta_p(m, r, w-1) \tag{29}$$
$$\vdots \tag{30}$$
$$|f_0 - f_1 + f_2 - \cdots + f_{p-1}|_{>w-(p-1)} \geq \Delta_p(m, r^*, w-(p-1)) \geq \Delta_p(m, r, w-(p-1)). \tag{31}$$

  Above, we used the fact that $\Delta_p(m, r, w)$ is a decreasing function of r, and $r^* \leq r$.

  Then, using the generalized Pascal's rule, we have

$$|f|_{>w} \geq \sum_{i=0}^{p-1} \Delta_p(m, r, w-i) = \Delta_p(m+1, r, w). \tag{32}$$

- Case 2: More generally, if $f^{(m+1)}|_{x_{m+1}\to z_j} = 0$ for some $\{z_j\}$, for $j = 1, \ldots A$, then

$$f^{(m+1)} = (x_{m+1} - z_1)(x_{m+1} - z_2)\cdots(x_{m+1} - z_A)h^{(m+1)},$$

  with $h^{(m+1)} \neq 0$. Note that $h^{(m+1)}$ may still depend on x_{m+1}, but $\deg h^{(m+1)} = \min(r-A, (m+1)(p-1)) = r - A$. We count the size of the support of $h^{(m+1)}$ as in Case 1, but we should not count any points for which $x_{m+1} = z_j$, for any z_j. Then, we have

$$|f^{(m+1)}|_{>w} \geq \sum_{i=0,\ i\notin\{z_1,\ldots,z_A\}}^{p-1} \Delta_p(m, r-A,\ w-i) \geq \sum_{i=0}^{p-1-A} \Delta_p(m, r-A, w-i). \tag{33}$$

  Using Lemma 6, in Appendix B, we have

$$\sum_{i=0}^{p-1-A} \Delta_p(m, r-A, w-i) \geq \Delta_p(m+1, r, w). \tag{34}$$

  Therefore, again $|f|_{>w} \geq \Delta_p(m+1, r', w)$ as required.

□

Note that this proof also implies that the generator matrix for $PRM_p(r, m, w)$ is full rank, if $\Delta_p(m, r, w) > 0$. In those cases where $\Delta_p(m, r, w) = 0$, the punctured generator matrix is not of full rank, and this theorem does not determine the distance of the punctured code.

4.4 A family of quantum Reed-Muller codes

Theorem 5. *Puncturing $RM_p(r,m)$ codes according to Manhattan weight gives rise to a family of qudit (with prime dimension p) triorthogonal codes with parameters $[[\binom{m}{>w}_p, \binom{m}{\leq w}_p, \Delta_p(m,\tilde{r},w)]]_p$ with $\tilde{r} = m(p-1)-r-1$, $3r < m(p-1)$ and $w < m(p-1)-r$, if $\Delta_p(m,r,w) > 0$.*

Proof. We have already shown that n is given by $\binom{m}{>w}_p$. The number of puncture locations is $\binom{m}{\leq w}_p$; this is equal to k if the punctured triorthogonal matrix is of full rank, which is guaranteed by Theorem 4, if $\Delta_p(m,r,w) > 0$. From Theorem 3, the distance of the quantum triorthogonal code is $d(\mathcal{G}_0^\perp) = d(PRM_p(m(p-1)-r-1,m,w))$, which is given by Theorem 4. □

Let us study the overhead of this code in the asymptotic limit of large m when r is maximal. To facilitate the analysis, we choose $m = 3\alpha$, where α is a large positive integer. Then, $r = \alpha(p-1)-1$ is the largest value of r such that $3r < m(p-1)$. For this choice of parameters, the distance of our quantum code simplifies to

$$d = \binom{\alpha}{> w}_p, \tag{35}$$

and the yield parameter is given by,

$$\gamma = \frac{\log(n/k)}{\log(d)} = \frac{\log\left(\binom{3\alpha}{>w}_p\right) - \log\left(\binom{3\alpha}{\leq w}_p\right)}{\log\left(\binom{\alpha}{>w}_p\right)}. \tag{36}$$

Let us now study the limit of this expression when $\alpha \to \infty$, keeping the ratio $\frac{w}{(p-1)m} = t$ held fixed. The parameter t ranges from 0 to 1. The answers can be expressed in terms of $H_p(\theta) = \lim_{m\to\infty} \binom{m}{\theta m}_p$ as shown in Section 4.2 and Appendix A, and is

$$\gamma_0(\theta) = \lim_{m\to\infty} \frac{\log_p \binom{m}{>w}_p - \log_p \binom{m}{\leq w}_p}{\log_p \binom{m/3}{>w}_p} \to \begin{cases} 3(1-H_p(\theta)) & \theta < \frac{p-1}{6} \\ 3\frac{1-H_p(\theta)}{H_p(3\theta)} & \theta > \frac{p-1}{6}. \end{cases} \tag{37}$$

where $\theta = t(p-1)$. We extremize $\gamma(t)$ as a function of t to find the optimal yield parameter.

Table 1 gives numerical values of asymptotic $\gamma_0(p)$ with p and their corresponding $t_0 = \theta_0(p)/(p-1)$ that minimizes $\gamma_p(\theta)$. A sample computation is illustrated in Appendix A. We also show in Appendix A that $\gamma_0(p) \sim \frac{1}{\ln p}$ for large p.

One can also ask, what are the smallest codes arising from this construction that have $\gamma < 1$? These are are listed in Table 2 and their block sizes are plotted as a function of p in Figure 2. We see that the required block size decreases dramatically as p increases, until $p = 23$, when Reed-Solomon codes achieve sublogarithmic distillation. However, in the following section, we show that one can construct punctured Reed-Muller codes that give rise to triorthogonal codes with $\gamma < 1$ with much smaller block sizes, if one is willing to search for better puncture locations.

p	$\gamma_0(p)$	$t_0(p)$
2	0.67799	0.27063
3	0.63215	0.27369
5	0.55914	0.27868
7	0.50786	0.28230
11	0.44108	0.28720
13	0.41785	0.28896
17	0.38273	0.29169
19	0.36901	0.29278
23	0.34663	0.29459

Table 1: Asymptotic values of the yield parameter $\gamma_0(p)$.

p	m	w	**Code Parameters**	**Length**
2	58	14	$[[288215893050995568, 14483100716176, 21700]]_2$	$\sim 2^{58}$
3	32	16	$[[1852445880782154, 574308069687, 3510]]_3$	$\sim 2^{51}$
5	16	16	$[[152186472515, 401418110, 429]]_5$	$\sim 2^{37}$
7	13	20	$[[96448935471, 440074936, 231]]_7$	$\sim 2^{36}$
11	7	19	$[[18874416, 612755, 35]]_{11}$	$\sim 2^{24}$
13	7	23	$[[60848853, 1899664, 35]]_{13}$	$\sim 2^{26}$
17	4	17	$[[77540, 5981, 15]]_{17}$	$\sim 2^{16}$
19	4	19	$[[121470, 8851, 15]]_{19}$	$\sim 2^{17}$
23	1	5	$[[17, 6, 3]]_{23}$	$\sim 2^4$

Table 2: Smallest punctured Reed-Muller codes with $\gamma < 1$ arising from our construction for different values of the qudit dimensionality p. For $p \geq 23$, punctured Reed-Solomon codes [29] achieve $\gamma < 1$. Results for $p = 2$ are from [22].

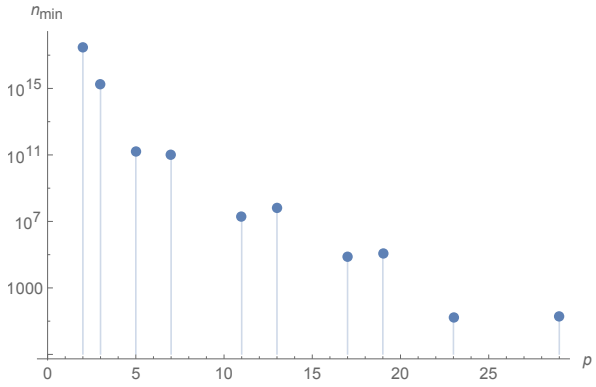

Figure 2: The minimum block size required for sublogarithmic distillation via our construction as a function of p. The code for $p = 2$ is from [22], and the vertical scale is logarithmic.

5 Sublogarithmic distillation with ~ 500 ququints

The puncturing scheme by Manhattan weight used in the earlier subsection was constructed by direct analogy from the scheme used in [22], and was chosen because it allowed for an explicit computation of the distance of the code for arbitrary p m and r, enabling us to compute γ in the limit of large n, as in [22]. However, there is no reason to expect that this scheme is the best way to puncture Reed-Muller codes. In this section, we attempt to ask, what are the best triorthogonal codes one can obtain from puncturing Reed-Muller codes?

This appears to be a non-trivial question to answer analytically and computationally; here, we present the best triorthogonal codes obtained from punctured Reed-Muller codes for small p, that we could find from a rudimentary randomized computer search. Similar searches over binary Reed-Muller codes appeared in [20]. We could very easily find codes

with values of k and/or d greater than or equal to what obtains from Manhattan distance puncturing. Some of the best triorthogonal codes we could find for $p = 3$ and $p = 5$ are listed in Table 3. Notably, we find a $[[519, 106, 5]]_5$ code with $\gamma = 0.99$. To our knowledge, no triorthogonal code with $\gamma < 1$ with block sizes less than 1000 for qubits or qutrits are known. We expect that randomly puncturing a ternary code Reed-Muller code with $m = 7$ would give rise to a triorthogonal codes with $\gamma < 1$, but confirming this was beyond the reach of our computational capabilities. The specific puncture locations that give rise to the codes in Table 3 are in Appendix C.

m	**Code Parameters**	A_d	γ
4	$[[80, 1, 5]]_3$	2080	2.72
4	$[[79, 2, 4]]_3$	130	2.65
4	$[[72, 9, 3]]_3$	648	1.89
5	$[[230, 13, 6]]_3$	572	1.60
5	$[[215, 28, 5]]_3$	1104	1.27
5	$[[206, 37, 4]]_3$	880	1.24
5	$[[200, 43, 3]]_3$	1700	1.40
6	$[[690, 39, 5]]_3$	1128	1.79
6	$[[667, 62, 4]]_3$	3972	1.71

m	**Code Parameters**	A_d	γ
2	$[[24, 1, 3]]_5$	96	2.89
2	$[[20, 5, 2]]_5$	760	2
3	$[[124, 1, 4]]_5$	124	3.48
3	$[[112, 13, 3]]_5$	512	1.96
4	$[[519, 106, 5]]_5$	2180	0.99

Table 3: Parameters of some triorthogonal codes obtained from punctured Reed-Muller spaces $RM_p(r_{\max}, m)$ for $p = 3$ (left) and $p = 5$ (right). These codes were found by searching over randomly chosen puncture locations, and are not claimed to be optimal. We find subloga-rithmic distillation at $m = 4$ for $p = 5$.

In the above table, we included the quantity A_d, which denotes the total number of weight-d logical-Z operators. A_d can be used to estimate the distillation performance of these $[[n, k, d]]_p$ codes in practice, when only a few rounds (or even only one round) of distillation is required. As shown in [4, 36], for distillation with triorthogonal codes, given n input qudits with depolarizing noise δ_{in}, we obtain k output qudits, with depolarizing noise

$$\delta_{\text{out}}^{(i)} = \frac{A_d^{(i)}}{(p-1)p^{d-1}} \delta_{\text{in}}^d, \tag{38}$$

for $i = 1, \ldots k$, where $A_d^{(i)}$ is the number of logical Z_i operators (i.e., logical Z operators on the ith output qudit) of weight d. It is slightly tedious to compute $A_d^{(i)}$ for each output qudit – we therefore estimate the average error of an output qudit to be, following [20], to be

$$\delta_{\text{out}} \approx \frac{A_d}{\bar{n}_T (p-1)p^{d-1}} \delta_{\text{in}}^d \left(1 - \frac{p-1}{p}\delta_{\text{in}}\right)^{n-d} \tag{39}$$

where A_d is the total number of weight-d logical-Z operators, and $\bar{n}_T \approx (1 - \frac{p-1}{p}\delta_{\text{in}})^n k$ is the average number of output magic-states. Applying this formula, we see that one round of distillation with the $[[519, 106, 5]]_5$ code, with $\delta_{\text{in}} = 10^{-3}$, we obtain $\delta_{\text{out}} \approx 8 \times 10^{-18}$, for a distillation cost $C = \frac{n}{\bar{n}_T} \approx 7.4$.

In our computational search, we did not seek to optimize A_d, only the yield parameter γ. To search for better distillation schemes, in situations where only one round of distillation is required, one should instead specify both δ_{in} and δ_{out}, and perform a computational search for puncture locations that optimize the distillation cost, C, rather than the yield.

6 Discussion

While there are difficulties associated with control of qudits, many have argued that there may be some advantages to working with qudits of larger dimension [8, 7, 29] particularly with respect to fault-tolerance, and several experimental groups are working towards realizing qudit based quantum computers (e.g., [30, 2, 46, 28, 39, 18, 31, 19, 42]). Quantifying and better understanding these potential advantages is important, as one explores all possible routes to fault-tolerant quantum computation.

Punctured Reed-Muller codes provide a natural construction of triorthogonal codes for magic state distillation, that can be readily applied to qudits of all prime dimensions. At least for this particular family of codes, we are able to quantify the advantages associated with increasing qudit dimensionality rather unambiguously via the yield parameter, γ that characterizes overhead of magic state distillation protocol. In particular, we find that, the puncturing scheme of [22], can be naturally generalized to qudits, and gives rise to asymptotic yield parameters in the limit $n \to \infty$ that decrease monotonically with increasing p, as shown in Table 1. At a more practical level, the smallest block size n required to obtain a sublogarithmic overhead, characterized by $\gamma < 1$, decreases rapidly with increasing p, from $p = 2$ up to $p = 23$, as shown in Table 2. These results are, of course, specific to Reed-Muller codes, and somewhat recently, better codes for magic state distillation have been proposed in the literature. Nevertheless, we believe that the results do illustrate and quantify the extent of the possible advantages for fault-tolerant overheads as one increases the qudit-dimensionality.

In experimental realizations, one may wish to focus on smaller values of p. For qutrits, the advantage is not as dramatic as for larger values of p, with $\sim 2^{52}$ qutrits required for sublogarithmic distillation compared to 2^{56} qubits [22], using our particular puncturing scheme. The codes in Table 2 are certainly not optimal, however, and better choices of puncture locations could lead to vastly superior codes. In a randomized search, we could not find any punctured Reed-Muller code for qutrits with $\gamma < 1$ for $n < 729$, although our search was not exhaustive. Therefore the, simple block construction of [36] still appears to be optimal for codes with less than 1000 qutrits. For ququints, however, we could find a set of puncture locations giving rise to a triorthogonal code with parameters $[[520, 105, 5]]_5$ with $\gamma < 1$.

The existence of these sublogarithmic distillation protocols, as well as the $[[520, 105, 5]]_5$ triorthogonal code, demonstrate that the space of qudit error-correcting codes is still relatively unexplored. It would be interesting to carry out an exhaustive search for small triorthogonal codes, like that of [33], for qudits, however, this is computationally challenging. We expect, however, that randomized searches and educated guesses could give rise to interesting codes.

A particularly interesting open question is to identify the smallest triorthogonal codes, for small values of p, that give rise to sublogarithmic distillation.

Acknowledgements

TS thanks Prof. Prasad Krishnan for fruitful discussions pertaining to the proof of the distance formula, Michael Zurel for his guidance and fruitful discussions. SP thanks Rev. Prof. Prem Saran Satsangi for inspiration and guidance. SP also thanks Prof. Nadish de Silva and Simon Fraser University for hospitality where part of this work was completed. Both authors acknowledge the support of DST-SERB grant (CRG/2021/009137).

Appendices

A Asymptotic analysis

In this section we work out asymptotic expressions for the p-nomial coefficients $\binom{m}{w}_p$, in the limit $m \to \infty$, with the ratio $w/m = \theta$ held fixed. We then use these expressions to determine the yield parameter $\gamma_\infty(p)$ in the asymptotic limit $m \to \infty$. We find that, in the limit of large p, $\gamma_\infty(p) \sim \frac{1}{\ln p}$.

A.1 Asymptotic expansions for p-nomial coefficients

We can write the p-nomial coefficient as a contour integral, directly from its definition,

$$\begin{aligned}\binom{m}{w}_p &= \frac{1}{w!}\frac{d^w}{dx^w}|_{x=0}(1+x+x^2+\dots x^{p-1}) \\ &= \frac{1}{2\pi i}\oint \frac{(1-z^p)^m}{(1-z)^m z^{w+1}}dz.\end{aligned} \tag{40}$$

The integration contour may be any closed contour, counter-clockwise, enclosing the origin. Now let $w = \theta m$, and assume $m \gg 1$. Note that

$$0 \leq \theta \leq (p-1). \tag{41}$$

We then have

$$\begin{aligned}\binom{m}{\theta m}_p =& \frac{1}{2\pi i}\oint \left(\frac{(1-z^p)}{(1-z)z^{\theta+\frac{1}{m}}}\right)^m dz. \\ =& \frac{1}{2\pi i}\oint e^{mf(z)}dz,\end{aligned} \tag{42}$$

where,

$$f(z) = \ln\left(\frac{(1-z^p)}{(1-z)z^{\theta+1/m}}\right) \approx \ln(1-z^p) - \ln(1-z) - \theta \ln z. \tag{43}$$

We evaluate this, in the limit of large m, by the method of steepest descent. Expressing our answers in terms of w rather than m, the result is

$$\binom{w/\theta}{w}_p \sim \frac{\xi^{-w-1}\left(\frac{\xi^p-1}{\xi-1}\right)^{\frac{w}{\theta}}}{\sqrt{2\pi}\sqrt{\frac{\theta+w\left(\theta+\frac{\xi^2}{(\xi-1)^2}-\frac{p(\xi^p+p-1)\xi^p}{(\xi^p-1)^2}\right)}{\theta\xi^2}}}. \tag{44}$$

where $\xi(\theta)$ is the dominant solution of the saddle-point equation, $f'(z)|_{z=\xi} = 0$, which is defined implicitly as the solution to the equation,

$$\frac{(p(\xi-1)-\xi)\xi^p+\xi}{(\xi-1)\,(\xi^p-1)} = \theta. \tag{45}$$

Defining,

$$H_p(\theta) = \lim_{m\to\infty}\frac{1}{m}\log_p\binom{m}{\theta m}_p = \lim_{w\to\infty}\frac{\theta}{w}\log_p\binom{w/\theta}{w}_p, \tag{46}$$

we have,

$$H_p(\theta) = \log_p\left(\frac{\xi^p-1}{\xi-1}\right) - \frac{\theta}{3}\log_p(\xi). \tag{47}$$

While this is a rather explicit expression for $H_p(\theta)$, one must solve a polynomial of degree $p-1$ to obtain $\xi(\theta)$. Let us consider a few small cases, to obtain explicit formulae.

For $p = 2$, the saddle-point equation is linear, and the dominant saddle is

$$\xi(\theta) = \frac{\theta}{1-\theta}, \tag{48}$$

so we find,

$$H_2(\theta) = -\log(1-\theta) - \theta\log\left(\frac{\theta}{1-\theta}\right). \tag{49}$$

This is, of course, just the binary entropy function.

For $p = 3$, the dominant saddle is

$$\xi = \frac{\theta+\sqrt{3(2-\theta)\theta+1}-1}{2(2-\theta)}. \tag{50}$$

We find,

$$H_3(\theta) = \log\left(\frac{7-3\theta+\sqrt{1-3\theta(\theta-2)}}{2(\theta-2)^2}\right) - \theta\log\left(\frac{\theta+\sqrt{1-3\theta(\theta-2)}-1}{2(2-\theta)}\right). \tag{51}$$

A plot of $H_3(\theta)$ compared with numerical values of $\binom{m}{w}_3$ for large m and w, is shown in Figure 3.

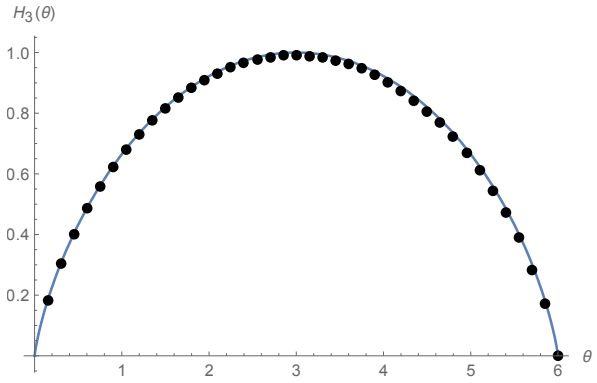

Figure 3: $H_3(2t)$ (blue) compared with numerical values (black points) computed for $m = \frac{w}{2t}$ and w, for $w = 400$.

For larger values of p, we find that the saddle point equation is solvable via computer algebra system, but the analytic solutions are rather unwieldy. Using these expressions, however, $H_p(\theta)$ can be computed numerically with ease. An alternative approach is to express H as a function of $\xi(\theta)$ rather than θ.

Note that the saddle-point equation determines $\theta(\xi)$, via a function that is monotonically increasing for $\xi \in (0, \infty)$. $\theta(0) = 0$ and $\lim_{\xi\to\infty} \theta(\xi) = p - 1$.

We therefore have the explicit expression,

$$H_p(\theta) = \log_p\left(\frac{\xi(\theta)^p - 1}{\xi(\theta) - 1}\right) + \frac{(\xi(\theta)^p(\xi(\theta)(1-p) + p) - \xi(\theta))}{(\xi(\theta) - 1)\,(\xi(\theta)^p - 1)} \log_p(\xi(\theta)), \tag{52}$$

where one may use $\xi(\theta) \in (0, \infty)$ instead of θ to parametrize the ratio of w and m.

A.2 Asymptotic expansion of p-nomial coefficients in the limit of large p

Let us now derive asymptotic formulae for large p. Let $\theta = t(p-1)$, so that $0 < t < 1$ for any p. We compute $\hat{H}_p(t) = H_p(t(p-1))$ when $p \to \infty$. The saddle point equation becomes,

$$(z^p - 1)\,(t(z-1) - z) - p(z-1)\,(t\,(z^p - 1) - z^p) = 0. \tag{53}$$

We use the ansatz that the dominant saddle point must take the form $\xi = 1 + \xi_1(t)/p + O(1/p^2)$, and will attempt to determine $\xi_1(t)$. This ansatz is consistent with numerical solutions of the saddle point equation at large p. We find $\xi_1(t)$ must satisfy the non-linear equation,

$$t = -\frac{1}{\xi_1} + \frac{1}{e^{\xi_1} - 1} + 1 \tag{54}$$

There is a unique solution $\xi_1(t)$ for $t \in (0, 1)$. In terms of $\xi_1(t)$ and t, we have,

$$\hat{H}_p(t) = \frac{t(p-1)}{w} \log_p \binom{w/t(p-1)}{w}_p \approx 1 - \frac{t\xi_1(t)}{\ln p} + \log_p\left(\frac{e^{\xi_1(t)-1}}{\xi_1(t)}\right). \tag{55}$$

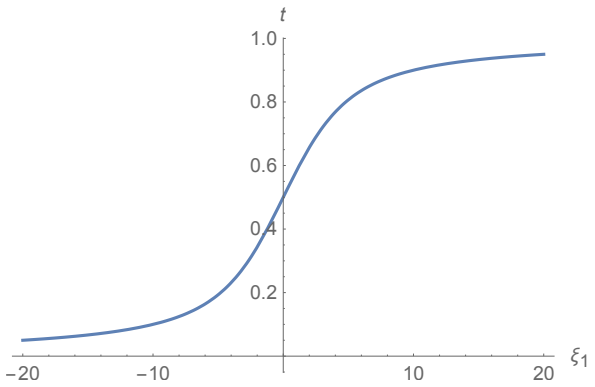

Figure 4: A plot of equation (54) for $t(\xi_1)$.

To proceed further we observe, as before, that $t(\xi_1)$ is given by Equation (54), which is monotonically increasing $t(-\infty) \to 0$ to $t(\infty) \to 1$, as shown in Figure 4. By choosing $\xi_1 \in (-\infty, \infty)$ we can parameterize all $t \in (0, 1)$. We now substitute in $t(\xi_1)$ using equation (54), to obtain,

$$\hat{H}_p(t) = 1 + \frac{1 - \frac{e^{\xi_1(t)}\xi_1(t)}{e^{\xi_1(t)}-1} + \ln\left(\frac{e^{\xi_1(t)}-1}{\xi_1(t)}\right)}{\ln(p)}. \tag{56}$$

A.3 Asymptotic limit of the yield parameter

We use the results of the previous section to work out the asymptotic limit of the yield parameter γ in the limit of $m \to \infty$ with $w/m = \theta$ is held fixed. Define,

$$H_p^{(\leq)}(\theta) = \frac{1}{m}\log\binom{m}{\leq \theta m}, \text{ and, } \quad H_p^{(>)}(\theta) = \frac{1}{m}\log\binom{m}{> \theta m}. \tag{57}$$

We clearly have

$$H_p^{(\leq)}(\theta) = \begin{cases} H_p(\theta) & \theta < (p-1)/2 \\ 1 & \theta > (p-1)/2 \end{cases} \tag{58}$$

and

$$H_p^{(>)}(\theta) = \begin{cases} 1 & \theta < (p-1)/2 \\ H_p(\theta) & \theta > (p-1)/2 \end{cases}. \tag{59}$$

The yield parameter, in the limit of large $m = 3\alpha$, is given by,

$$\gamma_0(\theta) = \lim_{m\to\infty} \frac{\log_p\binom{m}{>w}_p - \log_p\binom{m}{\leq w}_p}{\log_p\binom{m/3}{>w}_p} \to \begin{cases} 3(1 - H_p(\theta)) & \theta < \frac{p-1}{6} \\ 3\frac{1-H_p(\theta)}{H_p(3\theta)} & \theta > \frac{p-1}{6}. \end{cases} \tag{60}$$

In these expressions, we require $3\theta < p - 1$.

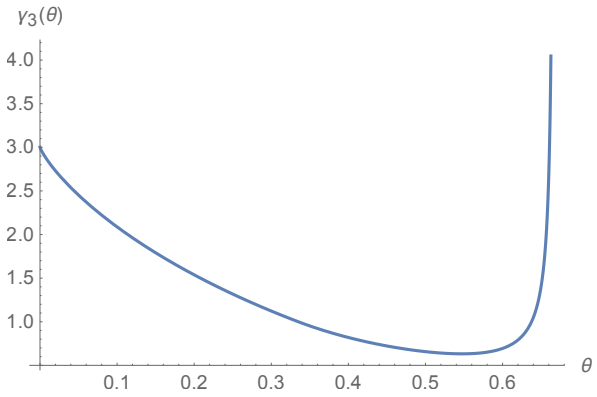

Figure 5: $\gamma(\theta)$ for $p = 3$. The minimum is at $\gamma = 0.632$, with $\theta = 0.547$

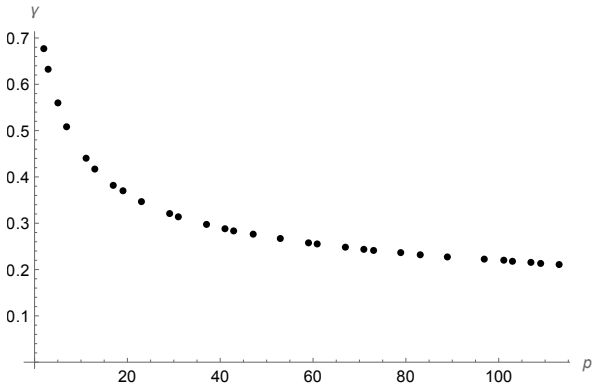

Figure 6: A plot of the best yield parameter γ as a function of p.

For $p = 3$, $\gamma_0(\theta)$ is plotted in Figure 5 and is given by,

$$\gamma_0(\theta) = \frac{3\left(-\theta\log(4-2\theta) + \log\left(\frac{1}{2}\left(-3\theta - \sqrt{1-3(\theta-2)\theta}+7\right)\right) + \theta\log\left(\theta + \sqrt{1-3(\theta-2)\theta}-1\right)\right)}{3\theta\log(4-6\theta) - 3\theta\log\left(3\theta+\sqrt{9\theta(2-3\theta)+1}-1\right) + \log\left(-\frac{6}{9\theta+\sqrt{9\theta(2-3\theta)+1}-7}\right)} \tag{61}$$

for $\theta > 1/3$. We find the minimum is $\gamma_0 = 0.632$, which occurs at $\theta_0 = 0.547$. Optimal yield parameters for other values of p are in Table 1 and plotted in Figure 6.

A.4 Asymptotic yield in the limit of large p

Let us consider the value of $\theta = (p-1)/6$, or $t = 1/6$. This is not claimed to be the optimal value of t, but it suffices to obtain a bound. (If t_* is the optimal value of t that minimizes γ, the we clearly have $\gamma_p(t_*) \le \gamma_p(1/6)$.) From equation (60), we have,

$$\gamma_p \to 3(1 - \hat{H}_p(t)). \tag{62}$$

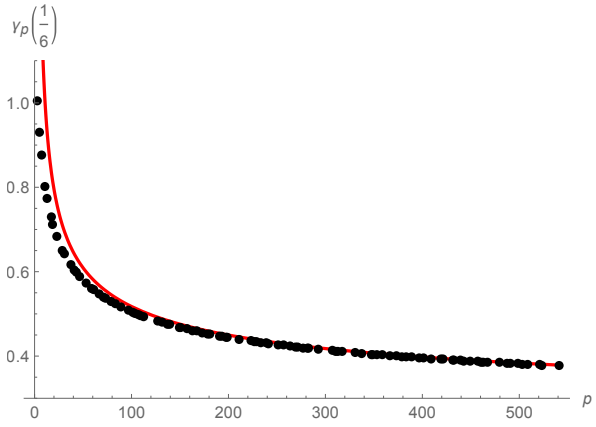

Figure 7: We plot the asymptotic yield parameter γ, evaluated at $t = \frac{w}{m(p-1)} = \frac{1}{6}$, from the asymptotic formula for large p, equation (63) (red line), and directly in the large m limit for finite p, (black points), for the first 100 primes. There is good agreement at larger p.

Using the expressions for $\hat{H}_p(t)$ in terms of $\xi_1(t)$ in Equation (56), we have,

$$\gamma_p(1/6) \to \frac{3\frac{e^{\xi_1(\frac{1}{6})}\xi_1(\frac{1}{6})}{e^{\xi_1(\frac{1}{6})}-1} - 3\ln\left(\frac{e^{\xi_1(\frac{1}{6})}-1}{\xi_1(\frac{1}{6})}\right)}{\ln p} \approx \frac{2.38309}{\ln p}. \tag{63}$$

Note that $\xi_1(t)$ is a constant independent of p. Thus $\gamma_p(1/6) \sim c/\ln p$ as $p \to \infty$. We evaluated the numerical constant c, using $\xi_1(1/6) \approx -5.903$. Equation (63) agrees well with numerics. We compare with the computation of $\gamma_p(1/6)$ exactly for the first 100 primes in Figure 7.

For *very* large p, $t_* \to \frac{1}{3}$, as, for any small positive value of ϵ, the denominator of $\frac{3(1-\hat{H}_p(1/3-\epsilon))}{\hat{H}_p(1-3\epsilon)} \to \frac{3(1-\hat{H}_p(1/3-\epsilon))}{1}$ for sufficiently large p. Thus $\gamma(t_*) \sim \frac{c}{\ln p}$, where $c \to 0.518$ from above, very slowly, as $p \to \infty$.

B Proof of technical lemma

We use the following lemma in the proof of Theorem 4.

Lemma 6.

$$\sum_{i=0}^{p-1-A} \Delta_p(m, r-A, w-i) \geq \sum_{i=0}^{p-1} \Delta_p(m, r, w-i) = \Delta_p(m+1, r, w). \tag{64}$$

with equality if $A = 0$.

Proof. Here $r = (p-1)\alpha + \beta$.

Case 1: $\beta \geq A$. We expand both sides and compare terms:

$$S_1 = \sum_{i=0}^{p-1-A} \Delta_p(m, r-A, w-i) = \sum_{i=0}^{p-1-A} \sum_{j=0}^{p-\beta+A-1} \binom{m-\alpha-1}{>w-i-j}_p \tag{65}$$

$$S_2 = \sum_{i=0}^{p-1} \Delta_p(m, r, w-i) = \sum_{i=0}^{p-1} \sum_{j=0}^{p-\beta-1} \binom{m-\alpha-1}{>w-i-j}_p \tag{66}$$

Note that, if $A = 0$, $S_1 = S_2$, so assume $A > 0$.

We see that,

$$\begin{aligned} S_1 - S_2 &= \sum_{i=0}^{p-1-A} \sum_{j=p-\beta}^{p-1-\beta+A} \binom{m-\alpha-1}{>w-(i+j)}_p - \sum_{i=p-A}^{p-1} \sum_{j=0}^{p-1-\beta} \binom{m-\alpha-1}{>w-(i+j)}_p \\ &= \sum_{i'=p-\beta}^{2p-\beta-A-1} \sum_{j'=0}^{A-1} \binom{m-\alpha-1}{>w-(i'+j')}_p - \sum_{i'=0}^{A-1} \sum_{j'=p-A}^{2p-1-\beta-A} \binom{m-\alpha-1}{>w-(i'+j')}_p \\ &= \sum_{i=p-\beta}^{p-A-1} \sum_{j=0}^{A-1} \binom{m-\alpha-1}{>w-(i+j)}_p \geq 0, \end{aligned} \tag{67}$$

and hence the theorem is proved. Here we used the fact that the summand only depends on $i + j$ in the second equality.

Case 2: $\beta < A$.

$$S_1 = \sum_{i=0}^{p-1-A} \Delta_p(m, r-A, w-i) = \sum_{i=0}^{p-1-A} \sum_{j=0}^{A-\beta} \binom{m-\alpha}{>w-i-j}_p \tag{68}$$

$$= \sum_{k=0}^{p-1} \sum_{i=0}^{p-1-A} \sum_{j=0}^{A-\beta} \binom{m-\alpha-1}{>w-i-j-k}_p . \tag{69}$$

S_2 is again given by eqn. (66). Notice the k summation in the S_1 is the same as the i summation on the S_2. It only remains to prove that

$$\sum_{i=0}^{p-1-A} \sum_{j=0}^{A-\beta} \binom{m-\alpha-1}{>w-i-j-k}_p \geq \sum_{j=0}^{p-1-\beta} \binom{m-\alpha-1}{>w-j-k}_p \quad \forall k.$$

It is easy to see that this is true – for any term corresponding to j on the RHS, we always have a corresponding term on the LHS that matches, and all terms are positive.

□

C Puncture locations

In this appendix, we list the puncture locations corresponding to the codes in section 5.

The unpunctured generator matrix for the triorthogonal space originating from a Reed-Muller code contains p^m columns. Our puncture locations are given as column numbers, where the first column is $c = 1$. Each column in the generator matrix for the triorthogonal space corresponds to a point in $\mathbb{F}_p^m$, $\vec{x} = (x_1, \ldots, x_m)$ with $x_i \in \mathbb{F}_p$. The cth column represents the point corresponding to $c - 1$ expressed in base-p, i.e., $c - 1 = x_1 + x_2 p + x_3 p^2 + \ldots$.

We only list puncture locations for codes with more than 2 punctures. For $p = 3$, we have:

- $[[72, 9, 3]]_3$:

  {12, 29, 34, 36, 53, 57, 63, 67, 75} (70)

- $[[230, 13, 6]]_3$:

  {43, 51, 74, 110, 140, 146, 147, 153, 180, 190, 198, 200, 228} (71)

- $[[215, 28, 5]]_3$:

  {4, 11, 38, 41, 43, 51, 52, 54, 57, 59, 67, 74, 77, 88, 110, 140, 146, 147, 153, 180, 183, 190, 191, 198, 200, 206, 222, 228 } (72)

- $[[206, 37, 4]]_3$:

  {5, 10, 13, 15, 23, 25, 37, 51, 55, 58, 62, 67, 78, 80, 85, 86, 93, 94, 95, 110, 113, 123, 137, 142, 146, 155, 166, 169, 180, 181, 182, 199, 203, 207, 208, 223, 228} (73)

- $[[200, 43, 3]]_3$:

  {5, 10, 13, 15, 23, 25, 27, 37, 51, 55, 58, 62, 67, 78, 80, 85, 86, 93, 94, 95, 110, 113, 114, 123, 137, 142, 146, 147, 155, 166, 169, 180, 181, 182, 199, 201, 203, 207, 208, 215, 218, 223, 228} (74)

- $[[690, 39, 5]]_3$:

  {32, 35, 46, 71, 75, 92, 117, 144, 146, 170, 181, 198, 205, 247, 255, 256, 261, 266, 277, 289, 303, 305, 330, 361, 372, 417, 424, 442, 477, 515, 527, 535, 557, 593, 605, 679, 713, 718, 729} (75)

- $[[667, 62, 4]]_3$:

  {9, 18, 32, 35, 46, 71, 75, 85, 92, 93, 117, 144, 146, 153, 166, 170, 174, 181, 198, 205, 238, 239, 247, 255, 256, 261, 266, 277, 289, 297, 303, 305, 312, 330, 347, 361, 371, 372, 417, 424, 442, 445, 477, 494, 515, 521, 522, 527, 535, 544, 557, 585, 593, 605, 648, 654, 656, 679, 713, 715, 718, 729} (76)

For $p = 5$

- $[[20, 5, 2]]_5$:

$$\{3,\ 6,\ 12,\ 16,\ 23\ \} \tag{77}$$

- $[[112, 13, 3]]_5$:

$$\{13,\ 18,\ 29,\ 33,\ 34,\ 46,\ 47,\ 58,\ 61,\ 79,\ 91,\ 111,\ 124\} \tag{78}$$

- $[[519, 106, 5]]_5$:

$$\begin{aligned}&\{9,\ 16,\ 19,\ 21,\ 44,\ 51,\ 54,\ 94,\ 99,\ 101,\ 106,\ 111,\ 117,\ 118,\ 119,\ 127,\ 129,\\ &137,\ 143,\ 144,\ 145,\ 146,\ 147,\ 148,\ 156,\ 169,\ 183,\ 197,\ 201,\ 204,\ 209,\ 220,\\ &223,\ 234,\ 241,\ 256,\ 268,\ 270,\ 273,\ 275,\ 280,\ 282,\ 285,\ 293,\ 294,\ 296,\ 314,\\ &321,\ 322,\ 325,\ 327,\ 341,\ 344,\ 353,\ 354,\ 357,\ 369,\ 376,\ 380,\ 408,\ 413,\ 423,\\ &443,\ 447,\ 448,\ 451,\ 452,\ 462,\ 472,\ 474,\ 476,\ 480,\ 482,\ 490,\ 493,\ 494,\ 498,\\ &504,\ 511,\ 512,\ 514,\ 519,\ 520,\ 524,\ 526,\ 529,\ 530,\ 532,\ 538,\ 545,\ 549,\ 552,\\ &556,\ 560,\ 562,\ 568,\ 570,\ 577,\ 590,\ 595,\ 597,\ 601,\ 606,\ 609,\ 611,\ 617\}\end{aligned} \tag{79}$$

References

[1] Hussain Anwar, Earl T Campbell, and Dan E Browne. Qutrit magic state distillation. *New J. Phys.*, 14(6):063006, June 2012. 1

[2] R. Bianchetti, S. Filipp, M. Baur, J. M. Fink, C. Lang, L. Steffen, M. Boissonneault, A. Blais, and A. Wallraff. Control and tomography of a three level superconducting artificial atom. *Phys. Rev. Lett.*, 105:223601, Nov 2010. 6

[3] Héctor Bombín, Mihir Pant, Sam Roberts, and Karthik I Seetharam. Fault-tolerant postselection for low-overhead magic state preparation. *PRX Quantum*, 5(1):010302, 2024. 1

[4] Sergey Bravyi and Jeongwan Haah. Magic-state distillation with low overhead. *Phys. Rev. A*, 86:052329, Nov 2012. 1, 2, 2, 5

[5] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Phys. Rev. A*, 71:022316, Feb 2005. 1, 3

[6] Earl T. Campbell. Enhanced fault-tolerant quantum computing in d-level systems. *Phys. Rev. Lett.*, 113:230501, Dec 2014. 1

[7] Earl T. Campbell. Enhanced fault-tolerant quantum computing in d-level systems. *Phys. Rev. Lett.*, 113:230501, Dec 2014. 1, 3.3, 3.3, 6

[8] Earl T. Campbell, Hussain Anwar, and Dan E. Browne. Magic-state distillation in all prime dimensions using quantum reed-muller codes. *Phys. Rev. X*, 2:041021, Dec 2012. 1, 2, 3, 3.2, 3.3, 3.3, 6

[9] Earl T Campbell and Dan E Browne. On the structure of protocols for magic state distillation. In *Workshop on Quantum Computation, Communication, and Cryptography*, pages 20–32. Springer, 2009. 1

[10] Earl T Campbell and Dan E Browne. Bound states for magic state distillation in fault-tolerant quantum computation. *Phys. Rev. Lett.*, 104(3):030503, 2010. 1

[11] Christopher Chamberland and Kyungjoo Noh. Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits. *npj Quantum Information*, 6(1):91, 2020. 1

[12] Hillary Dawkins and Mark Howard. Qutrit magic state distillation tight in some directions. *Phys. Rev. Lett.*, 115:030501, Jul 2015. 1

[13] Richard A. Demillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193–195, 1978. 3.1

[14] Leonhard Euler. De evolutione potestatis polynomialis cuiuscunque $1 + x + x^2 + x^3 + x^4 + \text{etc.}$. *Nova Acta Academiae Scientarum Imperialis Petropolitinae*, 12:47–57, 1801. Translated by Jordan Bell, `arXiv.math/0505425`. 4

[15] Austin G Fowler, Simon J Devitt, and Cody Jones. Surface code implementation of block code state distillation. *Scientific reports*, 3(1):1939, 2013. 3.3

[16] Craig Gidney, Noah Shutty, and Cody Jones. Magic state cultivation: growing t states as cheap as cnot gates, 2024. 1

[17] Louis Golowich and Venkatesan Guruswami. Asymptotically good quantum codes with transversal non-clifford gates. *arXiv preprint arXiv:2408.09254*, 2024. 1

[18] Noah Goss et al. High-fidelity qutrit entangling gates for superconducting circuits. *Nature Commun.*, 13(1):7481, 2022. [Erratum: Nature Commun. 14, 4256 (2023)]. 6

[19] Noah Goss, Samuele Ferracin, Akel Hashim, Arnaud Carignan-Dugas, John Mark Kreikebaum, Ravi K. Naik, David I. Santiago, and Irfan Siddiqi. Extending the computational reach of a superconducting qutrit processor, 2023. 6

[20] Jeongwan Haah and Matthew B. Hastings. Codes and Protocols for Distilling T, controlled-S, and Toffoli Gates. *Quantum*, 2:71, June 2018. 1, 3, 3.3, 5, 5

[21] Jeongwan Haah, Matthew B. Hastings, D. Poulin, and D. Wecker. Magic state distillation with low space overhead and optimal asymptotic input count. *Quantum*, 1:31, October 2017. 1

[22] Matthew B. Hastings and Jeongwan Haah. Distillation with sublogarithmic overhead. *Phys. Rev. Lett.*, 120:050504, Jan 2018. 1, 3, 3.3, 4.1, 4.1, 4.1, 4.2, 2, 2, 5, 6

[23] Mark Howard and Hillary Dawkins. Small codes for magic state distillation. *The European Physical Journal D*, 70:1–5, 2016. 1

[24] Mark Howard and Jiri Vala. Qudit versions of the qubit $\pi/8$ gate. *Phys. Rev. A*, 86:022316, Aug 2012. 1, 2

[25] Mark Howard, Joel Wallman, Victor Veitch, and Joseph Emerson. Contextuality supplies the 'magic' for quantum computation. *Nature*, 510(7505):351–355, June 2014. 1

[26] Akalank Jain and Shiroman Prakash. Qutrit and ququint magic states. *Phys. Rev. A*, 102:042409, Oct 2020. 1

[27] E. Knill. Fault-tolerant postselected quantum computation: Schemes, 2004. 1

[28] M. Kononenko, M. A. Yurtalan, S. Ren, J. Shi, S. Ashhab, and A. Lupascu. Characterization of control in a superconducting qutrit using randomized benchmarking, Oct 2021. 6

[29] Anirudh Krishna and Jean-Pierre Tillich. Towards low overhead magic state distillation. *Phys. Rev. Lett.*, 123:070507, Aug 2019. 1, 2, 2, 2, 6

[30] B. P. Lanyon, T. J. Weinhold, N. K. Langford, J. L. O'Brien, K. J. Resch, A. Gilchrist, and A. G. White. Manipulating biphotonic qutrits. *Phys. Rev. Lett.*, 100:060504, Feb 2008. 6

[31] Joseph Lindon, Arina Tashchilina, Logan W. Cooke, and Lindsay J. LeBlanc. Complete unitary qutrit control in ultracold atoms. *Phys. Rev. Appl.*, 19:034089, Mar 2023. 6

[32] Daniel Litinski. Magic state distillation: Not as costly as you think. *Quantum*, 3:205, 2019. 1

[33] Sepehr Nezami and Jeongwan Haah. Classification of small triorthogonal codes. *Phys. Rev. A*, 106:012437, Jul 2022. 1, 6

[34] Shiroman Prakash. Magic state distillation with the ternary Golay code. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476:20200187, 09 2020. 1

[35] Shiroman Prakash and Aashi Gupta. Contextual bound states for qudit magic state distillation. *Phys. Rev. A*, 101:010303, Jan 2020. 1

[36] Shiroman Prakash and Tanay Saha. Low overhead qutrit magic state distillation. *arXiv preprint arXiv:2403.06228*, 2024. 1, 5, 6

[37] Ben W. Reichardt. Quantum universality from magic states distillation applied to css codes. *Quantum Information Processing*, 4(3):251–264, August 2005. 1

[38] B.W. Reichardt. Quantum universality by state distillation. *Quantum Information and Computation*, 9(11 & 12):1030–1052, November 2009. 1

[39] Martin Ringbauer, Michael Meth, Lukas Postler, Roman Stricker, Rainer Blatt, Philipp Schindler, and Thomas Monz. A universal qudit quantum processor with trapped ions. *Nature Phys.*, 18(9):1053–1057, 2022. 6

[40] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, October 1980. 3.1

[41] Dayal Pyari Srivastava, Vishal Sahni, and Prem Saran Satsangi. From n-qubit multiparticle quantum teleportation modelling to n-qudit contextuality based quantum teleportation and beyond. *Int. J. Gen. Syst.*, 46(4):414–435, 2017. 1

[42] Mahadevan Subramanian and Adrian Lupascu. Efficient two-qutrit gates in superconducting circuits using parametric coupling, 2023. 6

[43] Wim van Dam and Mark Howard. Noise thresholds for higher-dimensional systems using the discrete Wigner function. *Phys. Rev. A*, 83:032310, Mar 2011. 1

[44] Eric W. Weisstein. Trinomial coefficient. From MathWorld–A Wolfram Web Resource. `https://mathworld.wolfram.com/TrinomialCoefficient.html`, 2023. Accessed: 2025-02-08. 4

[45] Adam Wills, Min-Hsiu Hsieh, and Hayata Yamasaki. Constant-overhead magic state distillation, 2024. 1

[46] M.A. Yurtalan, J. Shi, G.J.K. Flatt, and A. Lupascu. Characterization of multilevel dynamics and decoherence in a high-anharmonicity capacitively shunted flux circuit, Nov 2021. 6

[47] Richard Zippel. Probabilistic algorithms for sparse polynomials. In Edward W. Ng, editor, *Symbolic and Algebraic Computation*, pages 216–226, Berlin, Heidelberg, 1979. Springer Berlin Heidelberg. 3.1