E-PERFECT CODES

FRANCIS CASTRO, HEERALAL JANWA, GARY MULLEN,
AND IVELISSE RUBIO

ABSTRACT. Perfect codes provide one of the most important and widely studied classes of error-correcting codes. The problem of classifying the parameters of perfect codes remained open until 1973. In this paper we generalize perfect codes to e-perfect codes over finite fields. We present a list of parameters for e-perfect codes and conjecture that the list is complete, provide constructions for almost all the e-perfect codes listed in the conjecture and present partial results on proving that there are no other parameters for e-perfect codes.

1. INTRODUCTION

Let F_q denote the finite field of order q, where $q = p^r$ is a prime power. We will denote a linear code C over F_q of length n, dimension k, and minimum distance d as an $[n, k, d]$ code. The code C is a k-dimensional subspace (and thus C contains q^k codewords) of the vector space F_q^n. If a code C has length n, has M codewords and minimum distance d, but may not be linear, we will denote C as an (n, M, d) code. It is known that a code C can correct t errors if $t = \lfloor \frac{d-1}{2} \rfloor$. We assume that our codes do not have any zero columns; i.e., there is no coordinate in which every codeword has a zero in that coordinate.

A *sphere* $S_\rho(\mathbf{x})$ *of radius* ρ with center at the vector $\mathbf{x} = (x_1, \ldots, x_n) \in F_q^n$ is defined to be the subset $\{\mathbf{y} \in F_q^n : d(\mathbf{x}, \mathbf{y}) \leq \rho\}$, where $d(\mathbf{x}, \mathbf{y})$ denotes the Hamming distance (the number of coordinates in which $\mathbf{x}$ and $\mathbf{y}$ differ) between the vectors $\mathbf{x}$ and $\mathbf{y}$. The well known Hamming bound can be stated as:

Theorem 1.1. *Let C be a t-error-correcting code of length n over F_q. Then,*

$$|C| \left[1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t \right] \leq q^n.$$

A code C is said to be *perfect* if the code's parameters yield an equality in the Hamming bound. A code is thus perfect if for every possible vector $\mathbf{v}$ of length n, there is a unique codeword $\mathbf{c}$ in C in which at most t letters

of $\mathbf{c}$ differ from $\mathbf{v}$. The parameters of all perfect codes are known, and can be listed as follows:

The *trivial* perfect codes are the zero vector $(0,\dots,0)$ of length n, the entire vector space F_q^n, and the binary repetition code of odd length n. The non-trivial perfect codes must have the parameters $\big(n, M = q^k, 3\big)$ of the Hamming codes and the Golay codes (unique up to equivalence) whose parameters can be listed as follows [6, Theorem 6.33] (see [9, 12]).

1. The Hamming code $\left[\frac{q^m-1}{q-1}, n-m, 3\right]$ over F_q, where $m \geq 2$ is a positive integer;
2. The $[11, 6, 5]$ Golay code over F_3;
3. The $[23, 12, 7]$ Golay code over F_2.

Corollaries 20.16 and 20.21 in [6] imply that the Golay codes are the only perfect codes with parameters $\big(11, 3^6, 5\big)$ and $\big(23, 2^{12}, 7\big)$. However, there are other nonlinear single error correcting perfect codes, see [6, p. 180]. The reader can check that $n = 90, k = 78, t = 2$ and $q = 2$ yield an equality in the Hamming bound, but there is no code with these parameters. Research Problem 6.7 in [6] asks whether there exists a [90,77,5] binary code (see also [10, 11, 12]).

In [6, p. 187] there is a brief discussion and numerous references for perfect codes using the Lee metric, over mixed alphabets, and in graphs. See also [1] for a more recent paper discussing diameter perfect codes.

We now consider a generalization of perfect codes that includes codes with the above parameters. How close can we come to satisfying the Hamming bound without the code being perfect? The following provides one such measure. Let $0 \leq e \leq n$ be an integer. A t-error correcting code C with parameters $(n, M, d), t = \lfloor \frac{d-1}{2} \rfloor$, is said to be *e-perfect* if in the Hamming bound, equality is achieved when, on the right hand side, q^n is replaced by q^e. Thus an n-perfect code is of course a perfect code. We do not assume that an e-perfect code has to be linear. If an e-perfect code exists, then $|C| = q^k$ (the proof is similar to Lemma 3.4 in [6]). Therefore, to determine the existence of e-perfect codes one first has to determine q, n and t such that the equation

$$\sum_{i=0}^{t} \binom{n}{i} (q-1)^i = q^{e-k} = q^f \tag{1}$$

has positive integer solutions for $k \leq e \leq n$. Then we must see if there are codes with parameters $(n, M = q^k, d = 2t+1)$. A remark on page 184 of [6] tells us that the only solutions to (1) for $n, t \leq 1,000$ and $q \leq 100$ are the trivial solutions, the parameters of the Hamming and Golay codes, and $n = 90,\ k = 78,\ t = 2,\ q = 2$. Our search for $n \leq 10^4$, $t \leq 10^3$, and $q \leq 100$ yielded no other solutions. A much wider search, using the p-adic

root finding method of McAndrew [7] seems to indicate no other solutions.

We note that, for $q = 2$, Equation (1) always has a solution that corresponds to the trivial code.

Lemma 1.2. *The equation* $\sum_{i=0}^{t} \binom{n}{i} = 2^f$ *has solution* $n = 2t+1, f = 2t$.

Proof. We note that $\sum_{i=0}^{2t+1} \binom{2t+1}{i} = 2^{2t+1}$. We see that the first $t+1$ terms of the sum are equal to the last $t+1$ terms of the sum. Hence $\sum_{i=0}^{t} \binom{n}{i} = \frac{2^{2t+1}}{2} = 2^{2t}$. □

We remind the reader that if a code is linear, then F_q will be understood to mean the finite field containing q elements. However, if the code is non-linear we will simply view F_q as a set of q distinct symbols (without the usual algebraic structure of a finite field).

A simple computation shows that

1. for $t = 1$ and $n = \frac{q^m-1}{q-1}$, one has $M = q^k = q^{e-m}$;
2. for $q = 3, t = 2$ and $n = 11$, $M = q^k = 3^{e-5}$;
3. for $q = 2, t = 3$ and $n = 23$, $M = q^k = 2^{e-11}$;
4. and for $q = 2, t = 2$ and $n = 90$, $M = q^k = 2^{e-12}$.

We note that there are no binary perfect codes with $n = 90, t = 2$ [6, Theorem 6.38].

We now conjecture that the following is the complete list of parameters for all e-perfect codes over an alphabet with a prime power q number of symbols.

Conjecture 1.3. *Let C be an (n, M, d) non-trivial e-perfect code over F_q. Then C must have one of the following sets of parameters:*

1. $\left(\frac{q^m-1}{q-1}, q^{e-m}, 3\right)$, *with q a prime power and $m < e \leq n$, where $m \geq 2$ is an integer;*
2. $(11, 3^{e-5}, 5)$, *with $q = 3$ and $5 < e \leq 11$;*
3. $(23, 2^{e-11}, 7)$, *with $q = 2$ and $11 < e \leq 23$;*
4. $(90, 2^{e-12}, 5)$, *with $q = 2$ and $12 < e \leq 89$.*

In the next section we construct codes with each of the parameters listed above, except for the case when $n = 90$ and $e = 89$. As was the case for perfect codes, there could be many e-perfect codes with a given set of parameters. In Section 3 we provide partial results on proving that the above list is complete.

2. Existence of e-perfect codes

In this section we provide constructions for all of the e-perfect codes listed in Conjecture 1.3 except for the case when $n = 90, e = 89$. We

remark that, when constructing subcodes, we have to maintain the same parameters as the original code.

1. Let $t = 1$, $m \geq 2$ be a positive integer, and let q be a prime power, $n = \frac{q^m-1}{q-1}$. For $e = n$, there exist linear codes with these parameters and they have to be the Hamming perfect codes (see Problem 1.28 of [6]). There are also non-linear codes with these parameters (see page 180 of [6]).

   We can construct other e-perfect 1-error-correcting codes for $m < e < n$ just by taking subcodes of dimension $e - m$. The parity check matrix H of the Hamming code C with parameters $[n, k, d] = [\frac{q^m-1}{q-1}, n-m, 3]$ can be constructed by taking as columns the nonzero vectors in F_q^m whose first nonzero entry is 1. If $H = (A \mid I_m)$, then $G = \left(I_k \mid -A^T\right)$ is a $k \times n$ matrix and there is a row in G (and hence a codeword in C) with weight 3. If we delete s rows of G and leave at least one row of weight 3, we obtain a generator matrix for a code of dimension $k' = k - s$ and minimum distance 3. Thus, deleting $s = n - e$ rows of G we obtain a generator matrix for a code with parameters $\left[n = \frac{q^m-1}{q-1}, k' = e - m, d = 3\right]$. To get codes with full support, we need to make sure that the vector 1^n always belongs to the subcode.
2. The $(11, 3^6, 5)$ ternary code has to be the linear unique perfect ternary Golay code ([6], Corollary 20.21). For $e < 11$, e-perfect codes can be obtained from the Golay code $[11, 6, 5]$. There are rows of weight 5 in the generator matrix of the Golay code; by deleting other rows one can obtain subcodes of the desired dimension (any subcode containing the vector 1^{23} will have full support).
3. The $(23, 2^{12}, 7)$ binary code has to be the linear unique perfect binary Golay code ([6], Corollary 20.16). For $e < 23$, e-perfect linear codes exist as subcodes of the Golay code that contain the vector 1^{23}.
4. There is no perfect binary code with parameters $(90, 2^{78}, 5)$. Is there a 2 error-correcting e-perfect code? Yes. For example, we can construct a $[90, 76, 5]$ linear binary 88-perfect code from the $[127, 113, 5]$ binary BCH code. First, we set 38 information bits equal to 0. This yields a $[127, 75, 5]$ code C_1. Since the BCH code contains the vector $v := (1^{127})$, we then obtain a $[127, 76, 5]$ code $C_2 := \langle C_1, v\rangle$. We puncture C_2 in 37 places from the coordinates we originally selected, and obtain a [90,76,5] code with full support, yielding a binary 88-perfect code with the given parameters. This is a modification of the shortening, expurgation, augmentation, and puncturing processes described in Berlekamp [2].

Using a similar process, we can derive binary e-perfect codes with parameters $\left(90, 2^{e-12}, 5\right)$, $12 < e \leq 88$ from the $[127, 113, 5]$ binary BCH code.

Open questions:

1. Find the number of non-equivalent e-perfect codes with parameters $(11, 3^{e-5}, 5)$ and $q = 3$; $(23, 2^{e-11}, 7)$ and $q = 2$.
2. For the case $n = 90, t = 2$, can we do better, i.e., can we obtain a $(90, 2^{77}, 5)$ binary 89-perfect code? This is Research Problem 6.7 in [6].

3. There are no other parameters for e-perfect codes

We note that our problem to classify all e-perfect codes will be a very difficult project as the classification of e-perfect codes must include, as a special case, a classification of perfect codes. The existence of perfect codes is related to the existence of integral roots of the Lloyd polynomial. Can we relate the existence of e-perfect codes to a similar problem?

We now present some partial results in the classification of e-perfect codes. First, we discuss the solvability of Equation (1).

1. Let $t = 1$. Then Equation (1) becomes $1 + n(q-1) = q^f$. Therefore $n = \frac{q^f - 1}{q-1}$ and $M = q^{e-f}$, as predicted.
2. Let $t = 2$. Then Equation (1) becomes

$$1 + n(q-1) + \frac{n(n-1)}{2}(q-1)^2 = q^f. \tag{2}$$

Completing the square we obtain

$$2n(q-1) = q - 3 \pm \sqrt{8q^f + q^2 - 6q + 1}.$$

This implies that, for $t = 2$, Equation (1) only has solutions when

$$8q^f + q^2 - 6q + 1 = x^2 \tag{3}$$

for some x.

(a) Let $q = 2$. Then $x^2 = 2^{f+3} - 7$ and the only possible solutions are $x = 1, 3, 5, 11$ and 181 (see page 205 of [8]), corresponding to $n = 0, 1, 2, 5$ and 90. The first 3 codes cannot exist, for $n = 5, e = 5, k = 1$ and $t = 2$ we obtain the repetition code and for $n = 90$ we obtain $f + 3 = e - k + 3 = 15$. Therefore $k = e - 12$ as was predicted.

(b) Let $q = 3$. Then, Equation (2) becomes $n^2 = \frac{3^f - 1}{2}$. In [5] Ljunggren proved that the only non-trivial solution for this equation is $(n, f) = (11, 5)$. This solution gives parameters $(11, 3^{e-5}, 5)$, $5 < e \leq 11$ corresponding to the Golay code as predicted.

(c) Let $q = 5$. Then, completing the square, Equation (2) becomes $(4n-1)^2 = 2\left(5^f\right) - 1$. Setting $x = 4n-1$ we get the equation $x^2 = 2\left(5^f\right) - 1$. A computation and a result of Cohn in [3] show that the only solutions for this equation are $(x, f) = (7, 2)$ and $(3, 1)$. These solutions correspond to $(n, f) = (2, 2)$ and $(1, 1)$ respectively. Hence there are no e-perfect codes with $t = 2, q = 5$.

(d) Let $q = 7$. Then, $x^2 = 8q^f + q^2 - 6q + 1 = 2(7^f) + 2 = 8\left(7^f + 1\right) = \frac{64(7^f+1)}{7+1}$, and solving this equation is the same as solving an equation of the form $y^2 = \frac{p^f+1}{p+1}$. Le proved in [4] that this does not have a solution when f is odd. This implies that there are no e-perfect codes with parameters $\left(n, 7^k, d\right)$, where $f = e - k$ is odd.

3. Let $t = 3$ and $q = 2$. Then $n \geq 3$ and Equation (1) is equivalent to

$$(n+1)(n^2 - n + 6) = 3\left(2^{f+1}\right), f \geq 3. \tag{4}$$

Let $h = n + 1$. Then $h \geq 4$, $h\left(h^2 - 3h + 8\right) = 3\left(2^{f+1}\right)$, and $h = 3^v 2^s$, $v \in \{0, 1\}, s \geq 1$.

Suppose that $s \geq 4$. Then, $h^2 - 3h + 8 = 2^3\left(2^3 x_1 - 6x_2 + 1\right) = 3^{1-v}2^{f+1-s}$, and this implies that $f+1-s = 3$, $v \in \{0, 1\}$. Therefore $h^2 - 3h + 8 = h(h+3) + 8 = 8$ or 24, both contradictions.

If $s = 1, 2$ or 3, then the possible values for h are $4, 6, 8, 12$ and 24. It can be verified that the only solutions to (4) are $n = 3, 7$ and 23. For $n = t = 3$ there are no codes and for $n = 7$ just the trivial code. Hence the only possible non-trivial e-perfect codes with $q = 2, t = 3$ have parameters $(23, 2^{e-11}, 7), 11 < e \leq 23$.

4. Let $t = 5$ and $q = 2$. Then $n \geq 5$ and Equation (1) is equivalent to

$$\begin{aligned}(n+1)&\left(n^4 - 6n^3 + 31n^2 - 26n + 120\right)\\ &= 120\left(2^f\right) = (3)(5)\left(2^{f+3}\right). \end{aligned}\tag{5}$$

Let $h = n+1$. Then $h \geq 6$, $h\left(h^4 - 10h^3 + 55h^2 - 110h + 2^3(23)\right)$ $= 2^3(2x + 23) = (3)(5)\left(2^{f+3}\right)$, and $h = 3^v 5^u 2^s$, $v, u \in \{0, 1\}$.

Suppose that $s \geq 3$. Then, $h^4 - 10h^3 + 55h^2 - 110h + 184 = 3^{1-v}5^{1-u}2^{f+3-s}$. This implies that $2^3(2x + 23) = 3^{1-v}5^{1-u}2^{f+3-s}$, and $f + 3 - s = 3$. Therefore $h^4 - 10h^3 + 55h^2 - 110h + 184 = 3^{1-v}5^{1-u}8$, $v, u \in \{0, 1\}$. But the left hand side is larger than 120 for $h \geq 6$, and we obtain a contradiction.

If $s = 0, 1$ or 2, then the possible values for h are $6, 10, 12, 15, 20, 60$ and 120. The reader can verify that the only solutions to (5) are $n = 5$ and 11. There are no codes with $n = t = 5$; and $n = 11$ gives the trivial code. Hence there are no non-trivial e-perfect codes with $q = 2, t = 5$.

5. Let $t = 7, 9$ and $q = 2$. Using methods similar to those used for $t = 3$ and $t = 5$ we verified that there are no non-trivial e-perfect codes with $q = 2$, and $t = 7$ or 9.

The following table summarizes the possible parameters and existence of non-trivial e-perfect codes $(n, q^k, 2t + 1)$:

t	q	Possible parameters	Existence
1	p^r	$\left(\frac{q^m-1}{q-1}, q^{e-m}, 3\right)$, $m < e \leq n$	from the Hamming codes
2	2	$\left(90, 2^{e-12}, 5\right)$, $12 < e \leq 88$	from the $[127, 113, 5]$ BCH code
2		$(90, 2^{77}, 5)$	conjectured to exist
	3	$\left(11, 3^{e-5}, 5\right)$, $5 < e \leq 11$	from the $[11, 6, 5]$ Golay code
	5	none	none
	7	none for $e - k$ odd	none for $e - k$ odd
		unknown for $e - k$ even	unknown for $e - k$ even
	other	unknown	unknown
3	2	$\left(23, 2^{e-11}, 7\right)$, $11 < e \leq 23$	from the $[23, 12, 7]$ Golay code
	other	unknown	unknown
5, 7, 9	2	none	none
	other	unknown	unknown

Acknowledgements

The authors would like to thank Fernando Piñero for his help with algorithms and their MAPLE implementations, and the referee for the comments made to the original version of this paper.

References

[1] R. Ahlswede, H. K. Aydinian, and L. H. Khachatrian. On perfect codes and related concepts. *Des. Codes Cryptogr.*, 22(3):221–237, 2001.

[2] E. R. Berlekamp. *Algebraic coding theory*. Aegean Park Pr; Revised edition, 1984.

[3] J. H. E. Cohn. Perfect Pell powers. *Glasgow Math. J.*, 38(1):19–20, 1996.

[4] M. Le. On the Diophantine equation $(x^m + 1)(x^n + 1) = y^2$. *Acta Arith.*, 82(1):17–26, 1997.

[5] W. Ljunggren. Noen setninger om ubestemte likninger av formen $(x^n - 1)/(x - 1) = y^q$. *Norsk Mat. Tidsskr*, 25:17–20, (1943).

[6] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. II.* North-Holland Publishing Co., Amsterdam, 1977. North-Holland Mathematical Library, Vol. 16.

[7] M. H. McAndrew. An algorithm for solving a polynomic congruence, and its application to error-correcting codes. *Math. Comp.*, 19:68–72, 1965.

[8] L. J. Mordell. Note on the diophantine equation $ax^2 + by^2 + cz^2 + dt^2 = 0$. *Bull. Amer. Math. Soc.*, 38(4):277–282, 1932.

[9] G. L. Mullen and C. Mummert. *Finite fields and applications*, volume 41 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2007.

[10] A. Tietäväinen. On the nonexistence of perfect codes over finite fields. *SIAM J. Appl. Math.*, 24:88–96, 1973.

[11] J. H. van Lint. A survey of perfect codes. *Rocky Mountain J. Math.*, 5:199–224, 1975.

[12] J. H. van Lint. *Introduction to coding theory*, volume 86 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 1999.

Department of Mathematics, University of Puerto Rico, San Juan, PR 00931
E-mail address: franciscastr@gmail.com

Department of Mathematics, University of Puerto Rico, San Juan, PR 00931
E-mail address: hjanwa@gmail.com

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA
E-mail address: mullen@math.psu.edu

Department of Computer Science, University of Puerto Rico, Río Piedras, Box 23355, San Juan, PR 00931
E-mail address: iverubio@gmail.com