LE MATEMATICHE
Vol. LXIX (2014) – Fasc. I, pp. 275–284
doi: 10.4418/2014.69.1.21

CLASSIFICATION OF BINARY SYSTEMATIC CODES OF SMALL DEFECT

ALBERTO RAVAGNANI

In this paper non-trivial non-linear binary systematic AMDS codes are classified in terms of their weight distributions, employing only elementary techniques. In particular, we show that their length and minimum distance completely determine the weight distribution.

1. Introduction

Let q be a prime power and let $\mathbb{F}_q$ denote the finite field with q elements. A (non-linear) *code* of length $n \in \mathbb{N}_{\geq 1}$ over the field $\mathbb{F}_q$ is a subset $C \subseteq \mathbb{F}_q^n$ with at least two elements. We omit the adjective *non-linear* for the rest of the paper. A code $C \subseteq \mathbb{F}_q^n$ is said to be *linear* if it is a vector subspace of $\mathbb{F}_q^n$. Define the *Hamming distance* on $\mathbb{F}_q^n$ by $d : \mathbb{F}_q^n \times \mathbb{F}_q^n \to \mathbb{N}$ with

$$d(v,w) := |\{1 \leq i \leq n : v_i \neq w_i\}|,$$

where $v = (v_1, ..., v_n)$ and $w = (w_1, ..., w_n)$. The *minimum distance* $d(C)$ of a code $C \subseteq \mathbb{F}_q^n$ is the integer

$$d(C) := \min\{d(v,w) : v, w \in C, v \neq w\}.$$

A code of length n, $|C|$ codewords and minimum distance d is said to be of *parameters* $[n, |C|, d]$. The *weight* of a vector $v \in \mathbb{F}_q^n$ is the integer $\mathrm{wt}(v) := |\{1 \leq$

Entrato in redazione: 14 maggio 2013

AMS 2010 Subject Classification: 11T71, 68P30.
Keywords: Non-linear code, Systematic code, Binary code, AMDS code.

$i \leq n : v_i \neq 0\}|$, where $v = (v_1, ..., v_n)$. Let $C \subseteq \mathbb{F}_q^n$ be a code containing zero. For any $i \in \mathbb{N}$ such that $0 \leq i \leq n$ we denote by $W_i(C)$ the number of codewords in C having weight exactly i. The collection $\{W_i(C)\}_{0 \leq i \leq n}$ is said to be the *weight distribution* of C, and the $W_i(C)$'s are called *weights*. The following bound is well-known ([11], Theorem 1).

Proposition 1.1 (Singleton bound)**.** *Let $C \subseteq \mathbb{F}_q^n$ be a code of minimum distance d. Then $|C| \leq q^{n-d+1}$.*

Definition 1.2. A code $C \subseteq \mathbb{F}_q^n$ which attains the Singleton bound is said to be an *MDS* codes (MDS stands for *Maximum Distance Separable*). A code $C \subseteq \mathbb{F}_q^n$ with minimum distance d and cardinality q^{n-d} is said to be an *AMDS* code (AMDS stands for *Almost* MDS).

Hence, AMDS codes are codes that almost reach the Singleton bound. They have been introduced and studied for the first time in [1].

The remainder of the paper is organized as follows. Section 2 contains some preliminary results on the parameters of MDS and AMDS binary codes. We introduce systematic codes in Section 3, where we also classify binary systematic AMDS codes with respect to their parameters. In Section 4 we prove that length and minimum distance completely determine the weight distribution of such codes, and compute them explicitly.

2. Preliminaries

First, as an application of the well-known Hamming bound (see [9], Theorem 1.1.47), we prove powerful restrictions on the size of MDS and AMDS codes.

Proposition 2.1. *Let $C \subseteq \mathbb{F}_q^n$ be a code of minimum distance $d \geq 3$ and $|C| = q^k$ words.*

1. *If C is an MDS code, then $k \leq q - 1$.*
2. *If C is an AMDS code, then $k \leq q^2 + q - 2$.*

Proof. Set $s(C) := n - d - k + 1$, so that $s(C) = 0$ if C is MDS, and $s(C) = 1$ if C is AMDS. Remove from the codewords of C the last $d - 3$ components, obtaining a code, say D, of length $n - d + 3$, minimum distance at least 3, and $|C| = q^k$ codewords. Applying the Hamming bound to D we get

$$q^{n-d+1-s(C)} \left[1 + (n - d + 3)(q - 1)\right] \leq q^{n-d+3}.$$

Straightforward computations give the thesis. □

The following classification of binary MDS codes is a well-known result in coding theory (see for instance [7], Problem 5.32). Another proof using different techniques can be found in [5].

Theorem 2.2 (Classification of binary MDS codes)**.** *Let $C \subseteq \mathbb{F}_2^n$ be any MDS code of minimum distance d. Then, up to traslation, C is one of the following MDS codes.*

1. *The n-times repetition code, with $d = n \geq 3$.*
2. *The parity-check code of the code $\mathbb{F}_2^k$, $k = n-1$.*
3. *The code $\mathbb{F}_2^n$.*

Proposition 2.1 and Theorem 2.2 will be employed in the following sections to determine the possible parameters and weight distributions of binary systematic AMDS codes.

Definition 2.3. Codes C, D over a finite field $\mathbb{F}_q$ are said to be *P-equivalent* if they have the same parameters, i.e., $[n(C), |C|, d(C)] = [n(D), |D|, d(D)]$. Codes C, D over $\mathbb{F}_q$ and containing zero are said to be *W-equivalent* if they have the same length and the same weight distribution.

Remark 2.4. Notice that if C and D are linear W-equivalent codes, then they are also P-equivalent. For non-linear codes containing zero, this result is not true in general (see Example 2.5).

Example 2.5. The two binary codes

$$\{00000, 11001, 00111\}\ , \quad \{00000, 10011, 11001\}$$

contain the zero codeword, have the same weight distribution, and different minimum distances.

3. Parameters of binary systematic AMDS codes

Here we study binary systematic AMDS codes providing a classification in terms of their parameters. Let us briefly recall the definition of systematic code.

Definition 3.1. Let n be a positive integer and q a prime power. A code $C \subseteq \mathbb{F}_q^n$ is said to be *systematic* if there exists a function $\varphi : \mathbb{F}_q^k \to \mathbb{F}_q^{n-k}$ such that:

1. $\varphi(0) = 0$,
2. $C = \{(v, \varphi(v)) : v \in \mathbb{F}_q^k\}$.

The function φ is a *systematic encoding function*. A code C as in the definition has q^k codewords.

Remark 3.2. Notice that condition 1. is not always required in the definition of systematic code in the literature. On the other hand, up to a translation, we can always assume that 1. holds without loss of generality.

Systematic codes turn out to be very useful in the applications (see [10] and [2] among others) and powerful bounds on their parameters have been recently discovered (see e.g. [3]).

We study first binary systematic AMDS codes of minimum distance one and two, providing a characterization.

Proposition 3.3. *A code $C \subseteq \mathbb{F}_2^n$ of minimum distance $d = 1$ and 2^k codewords ($k \geq 1$) is systematic and AMDS if and only if there exists a function $\psi : \mathbb{F}_2^k \to \mathbb{F}_2$ with the following properties:*

(a) $\psi(0) = 0$,

(b) ψ *is not the parity-check function,*

(c) $C = \{(v, \psi(v)) : v \in \mathbb{F}_2^k\}$.

As a consequence, for any $n \geq 2$, there are $2^{2^{n-1}-1} - 1$ such codes.

Proof. Assume that C is systematic and AMDS. Let $\psi := \varphi$, the encoding function of Definition 3.1. We clearly have $\psi(0) = 0$. By contradiction, assume that φ is the parity-check function. Consider two vectors $v, w \in \mathbb{F}_2^k$ such that $d(v,w) = 1$. We have $\mathrm{wt}(v) \not\equiv \mathrm{wt}(w) \pmod 2$ and $d((v, \varphi(v)), (w, \varphi(w))) = 2$. This proves that the minimum distance of C is two, a contradiction. Now assume that $\psi : \mathbb{F}_2^k \to \mathbb{F}_2$ satisfies the hypothesis. We need to prove that the code $C := \{(v, \psi(v)) : v \in \mathbb{F}_2^k\}$ is AMDS. By contradiction, assume that C is not AMDS. Since $d(C)$ trivially satisfies $1 \leq d(C) \leq 2$, C has 2^k elements and length $k+1$, we have that C is an MDS code. This contradicts Theorem 2.2. □

Proposition 3.4. *The following facts hold.*

1. *A code $C \subseteq \mathbb{F}_2^n$ of minimum distance $d = 2$ and 2^k elements ($k \geq 2$) is systematic and AMDS if and only if there exists a function $\psi : \mathbb{F}_2^k \to \mathbb{F}_2^2$ such that:*

   (a) $\psi(0) = 0$,

   (b) $\psi(v) \neq \psi(w)$ *for any $v, w \in \mathbb{F}_2^k$ such that $d(v,w) = 1$,*

   (c) $C = \{(v, \psi(v)) : v \in \mathbb{F}_2^k\}$.

2. *A subset $C \subseteq \mathbb{F}_2^n$ is an AMDS systematic code with two elements if and only if it is of the form $\{0,(1,v)\}$ with $n \geq 2$, $v \in \mathbb{F}_2^{n-1}$ and $wt(v) = n-2$.*

Proof. If C is systematic and AMDS, let $\psi := \varphi$, the encoding function of Definition 3.1. Properties (a), (b) and (c) are easily checked. On the other hand, assume that $\psi : \mathbb{F}_2^k \to \mathbb{F}_2^2$ satisfies (a), (b), (c). By (a), the code $C := \{(v, \psi(v)) : v \in \mathbb{F}_2^k\}$ is systematic. By (b), the code C has not minimum distance one. By the Singleton bound, we have $d(C) \in \{2,3\}$. If C has minimum distance three, then it is an MDS code. On the other hand, Theorem 2.2 states that a binary MDS code of parameters $[k+2, 2^k, 3]$ does not exist ($k \geq 2$). Hence $d(C) = 2$ and C is AMDS. The last part of the claim is immediate. □

Now we focus on the P-classification of binary systematic AMDS codes with minimum distance at least three. We start by proving that the size of any such a code has to be very small.

Lemma 3.5. *Let C be a binary systematic AMDS code of minimum distance $d \geq 3$ and length n. Then $k := n - d \in \{1,2,3,4\}$. Moreover, if $k \in \{2,3,4\}$, then $d \in \{3,4\}$.*

Proof. Proposition 2.1 gives $k \in \{1,2,3,4\}$. If $k \in \{2,3\}$, then the Plotkin bound ([9], Theorem 1.1.45) implies $d \in \{3,4\}$. If $k = 4$, then the same bound gives $d \in \{3,4,5\}$. The case $k = 4$ and $d = 5$ is ruled out by the Hamming bound ([9], Theorem 1.1.47). □

Theorem 3.6 (P-classification)**.** *Let C be a binary systematic AMDS code of length n and minimum distance d. Then (n,d) is one of the following pairs:*

(a) $(n,1)$, *with* $n \geq 3$,
(b) $(n,2)$ *with* $n \geq 4$,
(c) $(d+1,d)$ *with* $d \geq 1$,
(d) $(5,3)$,
(e) $(6,4)$,
(f) $(6,3)$,
(g) $(7,4)$,
(h) $(7,3)$,
(i) $(8,4)$.

Moreover, for any such a pair (n,d) there exists a binary systematic AMDS code of length n and minimum distance d.

Proof. Assume that (n,d) are the length and the minimum distance of a binary systematic AMDS code. Combining Proposition 3.3, Proposition 3.4 and Lemma 3.5 we easily see that (n,d) must be one of the pairs in the list.

We need to show that, for any pair (n,d) in the list, there exists a binary systematic AMDS code with length n and minimum distance d. Proposition 3.3 and Proposition 3.4 produce examples of binary systematic AMDS codes with the parameters of (a), (b) and (c). Notice that, for any $n \geq 5$ and $d \geq 3$ odd, the parity-check code of a code with length n and minimum distance d has length $n+1$ and minimum distance $d+1$. As a consequence, it is enough to prove the theorem for the pairs $(5,3)$, $(6,3)$ and $(7,3)$. For each tern $[5,2^2,3]$, $[6,2^3,3]$ and $[7,2^4,3]$ we give in Table 1 a generator matrix of a binary linear systematic code having these parameters. We point out that the code of parameters $[7,2^4,3]$ in the table is the Hamming code $H(q=2,r=3)$ (see [8], page 23). □

$[n,2^k,d]$ $\rightarrow$	$[5,2^2,3]$	$[6,2^3,3]$	$[7,2^4,3]$
Generator matrix $\rightarrow$	$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$

Table 1: Binary systematic codes of parameters $[5,2^2,3]$, $[6,2^3,3]$ and $[7,2^4,3]$

4. Weight distributions of binary systematic AMDS codes

Here we focus on the W-classification of binary systematic AMDS codes of minimum distance at least three and more than two codewords. By Lemma 3.5, it is enough to study the weight distributions of codes of parameters $(n,d) \in \{(5,3),(6,4),(6,3),(7,4),(7,3),(8,4)\}$. We will treat the pairs $(5,3)$ and $(6,4)$ by using a computational approach (the computations take only a few seconds on a common laptop), and the other cases theoretically. The following lemma is proved by exhaustive research.

Lemma 4.1. *The weight distribution of any binary systematic AMDS code of length n and minimum distance d, with $(n,d) \in \{(5,3),(6,3)\}$, depends only on n and d, and it is given in Table 2. Moreover, any such a code is linear.*

Now we focus on the other pairs $(n,d) \in \{(6,4),(7,4),(7,3),(8,4)\}$ from a theoretical viewpoint. We notice that exhaustive search does not produce any result in a resonable time on a common computer when analyzing the cases $(n,d) = (7,3)$ and $(n,d) = (8,4)$. Let us first recall the definition of weight distribution of a code.

Values of (n,d)	Non-zero weights of any binary systematic code C with length n and minimum distance d
$(5,3)$	$W_0(C)=1, W_3(C)=2, W_4(C)=1$
$(6,3)$	$W_0(C)=1, W_3(C)=4, W_4(C)=3$

Table 2: Partial W-classification of binary systematic AMDS codes.

Definition 4.2. Let $C \subseteq \mathbb{F}_q^n$ be a code over a finite field $\mathbb{F}_q$. For any integer $i \in \{0,1,...,n\}$ define the integer $B_i(C)$ by $|C| \cdot B_i(C) := |\{(v,w) \in C^2 : d(v,w) = i\}|$. The collection $\{B_i(C)\}_{i=0}^n$ is called the *distance distribution* of C.

Remark 4.3. In the notation of Definition 4.2, if C is a linear code then its weight distribution and its distance distribution agree, i.e., $W_i(C) = B_i(C)$ for any $i \in \{0,1,...,n\}$.

Lemma 4.4. *A binary systematic AMDS code C of length $n = 7$ and minimum distance $d = 3$ has the following weight and distance distribution.*

$$\begin{array}{l|l} W_0(C)=1 & W_4(C)=7 \\ W_1(C)=0 & W_5(C)=0 \\ W_2(C)=0 & W_6(C)=0 \\ W_3(C)=7 & W_7(C)=1 \end{array} \qquad \begin{array}{l|l} B_0(C)=1 & B_4(C)=7 \\ B_1(C)=0 & B_5(C)=0 \\ B_2(C)=0 & B_6(C)=0 \\ B_3(C)=7 & B_7(C)=1 \end{array}$$

Proof. We clearly have $|C| = 2^4 = 16$, and so the parameters of C attain the Hamming bound ([9], Theorem 1.1.47). Such a code is said to be a perfect code (see [8], Chapter 6 and [4], Chapter 11). By [8], Theorem 37 at page 182 and the following remark, C has the same weight distribution of the well-known Hamming code $H(q=2,r=3)$ of parameters $[7,2^4,3]$ (see [8], pag. 23). The weight distribution of this simple linear code is well-known. □

The following result is immediate.

Lemma 4.5. *Let n be a positive integer and let $v,w \in \mathbb{F}_2^n$. Then $d(v,w) = wt(v) + wt(v) - 2v\cdot w$, where $v\cdot w = |\{1 \le i \le n : v_i = w_i = 1\}|$. In particular, the integer $wt(v) - wt(w)$ is odd if and only if $d(v,w)$ is odd.*

Theorem 4.6 (W-classification)**.** *Any binary systematic AMDS code of minimum distance at least three and cardinality at least four has exactly one of the weight distributions listed in Table 3. Moreover, each of those weight distribution corresponds to a binary systematic AMDS code.*

$[n,2^k,d] \to$	$[5,2^2,3]$	$[6,2^2,4]$	$[6,2^3,3]$	$[7,2^3,4]$	$[7,2^4,3]$	$[8,2^4,4]$
$W_0(C)$	1	1	1	1	1	1
$W_1(C)$	0	0	0	0	0	0
$W_2(C)$	0	0	0	0	0	0
$W_3(C)$	2	0	4	0	7	0
$W_4(C)$	1	3	3	7	7	14
$W_5(C)$	0	0	0	0	0	0
$W_6(C)$	-	0	0	0	0	0
$W_7(C)$	-	-	-	0	1	0
$W_8(C)$	-	-	-	-	-	1

Table 3: Weight distribution of any binary systematic AMDS codes C with minimum distance at least three and cardinality at least four.

Proof. Combining Theorem 3.6, Lemma 4.1 and Lemma 4.4, we see that it is enough to show that any binary systematic AMDS code of parameters $(n,d) \in \{(6,4),(7,4),(8,4)\}$ is the parity-check code of an AMDS systematic code of parameters $(n-1,d-1)$. Let C be a binary systematic AMDS code of parameters $(n,d) \in \{(6,4),(7,4),(8,4)\}$. Denote by E the code obtained by removing from the codewords of C the last component. Notice that E is either an MDS code, or a systematic AMDS code. By Theorem 2.2, the first case is ruled out. Hence E is a systematic AMDS code of parameters $(n-1,d-1) \in \{(5,3),(6,3),(7,3)\}$ (respectively). Clearly, there exists a function $f : E \to \mathbb{F}_2$ such that $C = \{(e, f(e) : e \in E)\}$. We will prove that f is the parity-check funtion on E, examining the three cases separately.

1. Assume $(n,d) = (6,4)$, so that E has length $n-1 = 5$ and minimum distance $d-1 = 3$. We clearly have $f(0) = 0$. By Lemma 4.1, E has two codewords of weight three and one of weight four. Let $w \in E$ of weight three. Since C has minimum distance 4, $f(w) = 1$. Let w' be the codeword of E of weight 4, and fix $w \in E$ of weight 3. By Lemma 4.5, we have $d(w',w) \in \{3,5\}$. If $d(w',w) = 5 = n-1$, then $w = (1,1,1,1,1) - w'$, which contradicts $\mathrm{wt}(w) = 4$. So $d(w',w) = 3$. Since C has minimum distance 4 and $f(w) = 1$, we must have $f(w') = 0$.

2. Assume $(n,d) = (7,4)$, so that E has length $n-1 = 6$ and minimum distance $d-1 = 3$. Again, $f(0) = 0$. By Lemma 4.1, E has four codewords of weight 3 and three of weight 4. Since C has minimum distance 4, we

have $f(w) = 1$ for any $w \in E$ of weight 3. Now fix a codeword $w \in E$ of weight 3 and let $w' \in E$ be any codeword of weight 4. By Lemma 4.5, we have $d(w', w) \in \{3, 5\}$. The case $d(w', w) = 5$ is ruled out by Remark 4.3. Indeed, Lemma 4.1 states that E is linear, and so its distance distribution agrees with its weigh distribution (given in Lemma 4.1). As a consequence, there are no codewords in E whose Hamming distance is five. Since C has minimum distance 4 and $f(w) = 1$, we must have $f(w') = 0$.

3. Assume $(n, d) = (8, 4)$, so that E has length $n - 1 = 7$ and minimum distance $d - 1 = 3$. We clearly have $f(0) = 0$. Since $d(C) = 4$, we get $f(w) = 1$ for any $w \in E$ of weight 3. Let $w' \in E$ be any codeword of weight 4 and $w \in E$ a fixed codeword of weight 3. By Lemma 4.5 and Lemma 4.4, we have $d(w', w) \in \{3, 7\}$. The case $d(w', w) = 7$ is easily ruled out. Since $f(w) = 1$ and C has minimum distance 4, we have $f(w') = 0$. Finally, fix a codeword $w' \in E$ of weight 4. We have $d(w', \underbrace{1111111}_{7}) = 3$. Since $f(w') = 0$, it is clear that $f(\underbrace{1111111}_{7}) = 1$.

□

Remark 4.7. We notice that the W-classification of Theorem 4.6 may be obtained also in the following way. Define an *isometry* on $\mathbb{F}_q^n$ as a map $i : \mathbb{F}_q^n \to \mathbb{F}_q^n$ preserving the Hamming distance between elements of $\mathbb{F}_q^n$. Codes $C, D \subseteq \mathbb{F}_q^n$ are said to be *isometric* if $D = i(C)$ for some isometry i. If we combine [6], Theorem 7.17 and [6], Table 7.2, we can easily see that for any pair $(n, d) \in \{(5,3), (6,4), (6,3), (7,4), (8,3), (8,4)\}$ there exists a unique, up to isometry, binary code of length n, minimum distance d, and 2^{n-d} codewords. Since isometric codes have the same distance distribution, we deduce that the W-classification of binary systematic AMDS codes must produce a unique equivalence class for each pair (n, d) in the list. As a consequence, it is enough to compute the weight distribution of just one code for each pair in order to get the whole W-classification. On the other hand, we notice that the proof here proposed uses elementary techniques, while the classification of [6] refers to non-trivial results of design and group theory.

Acknowledgment

The author would like to thank Emanuele Bellini, Elisa Gorla, Simon Litsyn and Massimiliano Sala for useful suggestions.

REFERENCES

[1] M. A. de Boer, *Almost MDS Codes*, Designs, Codes and Cryptography 9 (2) (1996), 143–155.

[2] R. D. Baker - J. H. van Lint - R. M. Wilson, *On the Preparata and Goethals codes*, IEEE Transactions on Information Theory 29 (1983), 342–345.

[3] E. Bellini - E. Guerrini - M. Sala, *A bound on the size of linear codes and systematic codes*, `http://arxiv.org/abs/1206.6006`.

[4] G. Cohen - I. Honkala - S. Litsyn - A. Lobstein, *Covering Codes*, North-Holland Mathematical Library 54, 1997.

[5] E. Guerrini - M. Sala, *A classification of MDS binary systematic codes*, BCR preprint, UCC Cork, Ireland, 2006.

[6] P. Kaski - P. R. J. Östergård, *Classification Algorithms for Codes and Designs*, Springer-Verlag, Berlin Heidelberg, 2006.

[7] S. Ling - C. Xing, *Coding Theory: A First Course*, Cambridge University Press, 2004.

[8] F. J. MacWilliams - N. J. A. Sloane, *The Theory of Error-Correcting codes*, North Holland Mathematical Library, 1977.

[9] D. Nogin - M. Tsfasman - S. Vlăduţ, *Algebraic Geometry Codes, Basic Notions*, American Mathematical Society, Series *Mathematical Surveys and Monographs*, 2007.

[10] F. Preparata, *A class of optimum nonlinear double-error correcting codes*, Information and Control 13 (1968), 378–400.

[11] R. C. Singleton, *Maximum distance q-ary codes*, IEEE Transactions on Information Theory 10 (1964), 116–118.

ALBERTO RAVAGNANI
Institut de Mathématiques
Université de Neuchâtel
Emile-Argand 11, CH-2000 Neuchâtel
Switzerland
e-mail: `alberto.ravagnani@unine.ch`