The weight spectrum of two families of Reed-Muller codes

Claude Carlet,*

Universities of Paris 8, France and Bergen, Norway. *E-mail*: claude.carlet@gmail.com

Patrick Solé,

I2M (CNRS, Aix-Marseille University, Centrale Marseille), Marseilles, France, E-mail: sole@enst.fr

Abstract

We determine the weight spectra of the Reed-Muller codes RM(m-3,m) for $m \ge 6$ and RM(m-4,m) for $m \ge 8$. The technique used is induction on m, using that the sum of two weights in RM(r-1,m-1)is a weight in RM(r,m), and using the characterization by Kasami and Tokura of the weights in RM(r,m) that lie between its minimum distance 2^{m-r} and the double of this minimum distance. We also derive the weights of RM(3,8), RM(4,9), by the same technique. We conclude with a conjecture on the weights of RM(m-c,m), where cis fixed and m is large enough.

Keywords: Reed Muller codes, weight spectrum MSC (2020): 94B27, 94D10

1 Introduction

For classical Coding Theory notation and terminology, see the next Section. Determining the Hamming weights in Reed-Muller codes has been considered

^{*}The research of the author is partly supported by the Norwegian Research Council.

an important research topic for more than half a century [6, Chapt. 15]. The weights of the Reed-Muller codes of length 2^m and orders 0, 1, 2, m-2, m-21, m are known (as well as their weight distributions). These weights equal $0, 2^m$ for the order 0, with additionally 2^{m-1} for the order 1, and $2^{m-1} \pm 2^i$ where $\frac{m}{2} \leq i \leq m$ for the order 2, see e.g. [6]. The weights in RM(m,m) are all integers between 0 and 2^m since $RM(m,m) = \mathbb{F}_2^{2^m}$. The weights in RM(m-1,m) are all even integers between 0 and 2^m and those in RM(m-1,m)(2, m) (the extended Hamming code) are all even integers between 0 and 2^{m} except 2 and $2^m - 2$: these two results (as well as the weight distributions) are directly deduced from the Mac Williams identity (see e.g. [6, Chapt. 5] or [3]), which says that the weight distribution of a linear code can be obtained as a function of the weight distribution of its dual; the duals of RM(m-1,m)and RM(m-2, m), namely RM(0, m) and RM(1, m), being such that $A_0 =$ $A_{2^m} = 1$, and for the latter, $A_{2^{m-1}} = 2^{m+1} - 2$, and all the other coefficients being zero, the weight distributions of RM(m-2,m) and RM(m-1,m) are easily deduced. Another method for determining the weight distribution of RM(m-2,m) is given in [11], by induction on m; we do not give it here since it is more complex than using the MacWilliams transform, but the idea of an induction will be central for the other orders that we shall address. Indeed, the weight distribution of RM(2,m) being already rather complex, because its coefficient $A_{2^{m-1}}$ has no known compact expression, it seems impossible to deduce the weight distribution of RM(m-3,m) from that of its dual, and this is still more true for RM(m-4,m).

All the low Hamming weights are known in all Reed-Muller codes: Berlekamp and Sloane [2] (see the Addendum in this paper) and Kasami and Tokura [4] have shown that, for $r \ge 2$, the only Hamming weights in RM(r, m) occurring in the range $[2^{m-r}; 2^{m-r+1}]$ are of the form $2^{m-r+1} - 2^i$ for some *i*; and the latter have completely characterized the codewords: the corresponding functions are affinely equivalent to

$$\begin{cases} x_1 \cdots x_{r-2} (x_{r-1} x_r \oplus x_{r+1} x_{r+2} \oplus \cdots \oplus x_{r+2l-3} x_{r+2l-2}), & \text{for } 2 \le 2l \le m-r+2, \\ x_1 \cdots x_{r-l} (x_{r-l+1} \cdots x_r \oplus x_{r+1} \cdots x_{r+l}), & \text{for } 3 \le l \le \min(r, m-r). \end{cases}$$

The functions whose Hamming weights are strictly less than 2.5 times the minimum distance $d = 2^{m-r}$ have later been studied in [5].

Those possible weights of the codewords in the Reed-Muller codes of orders $3, \ldots, m-4$ whose values lie between 2.5 d and $2^m - 2.5 d$ are unknown (but when m = 2r + 1, they are known in some cases by using invariant theory,

because the code is then self-dual, see [6, 10]).

In this note, we completely determine the weights in RM(m-3,m), and RM(m-4,m). Note that our work is on a different track than works such as [9] in which the authors consider some weights in some Reed-Muller codes and look for all the functions having these weights, whereas we are looking for the weights (all of them) in some Reed-Muller codes and we do not try to find the functions having these weights. We first observe that the work of [11, Theorem 16 (2)] can be made much more precise: this reference provides some weights in RM(m-3,m) and we show that these weights are in fact all weights thanks to the result of Kasami-Tokura [4]. We extend then the method to the codes RM(m-4,m).

The material is arranged as follows. The next section recalls some basic definitions and notions needed to understand the rest of the paper. Sections 3, and 4 study the weight spectra of RM(m-3,m), and RM(m-4,m), respectively. Section 5 describes some numerical examples. Section 6 concludes the article.

2 Preliminaries

The Hamming weight (in brief, the weight) of an element $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$ is the number of indices *i* such that $x_i \neq 0$. A binary linear code of length n is an \mathbb{F}_2 -subspace of \mathbb{F}_2^n . Its dimension is its dimension as an \mathbb{F}_2 -vector space. Its minimum distance (in brief, distance) is the minimum Hamming weight of a nonzero codeword.

The extension \widehat{C} of a linear code C of length n is the linear code of length n+1 such that each codeword $(c_0, \ldots, c_n) \in \widehat{C}$ satisfies both following conditions

1. $(c_0, \ldots, c_{n-1}) \in C$,

2.
$$\sum_{i=0}^{n+1} c_i = 0.$$

The weights of a code are the Hamming weights of all its codewords. The set of distinct weights (including the zero weight) is called the *weight* spectrum¹. The list of the numbers, classically denoted by A_i , of codewords

¹Contrary to the spectra used for instance in Boolean function theory, the weight spectrum in coding theory does not include the indication of the multiplicities of the weights.

of weight *i* for *i* ranging from 0 to *n* is called the *weight distribution* of the code. The Magma notation for this quantity is the list of pairs $\langle i, A_i \rangle$ where $A_i \neq 0$.

The *Reed Muller* codes are a family of binary linear codes of length $n = 2^m$. Given the order $r \in \{0, \ldots, m\}$ of such a code (usually denoted by RM(r,m)), the dimension equals $\sum_{i=0}^{r} {m \choose i}$ and the minimum distance equals 2^{m-r} . An explicit definition in terms of Boolean functions is as follows. Let B_m denote the vector space of polynomials in m variables with coefficients in \mathbb{F}_2 , that is, of elements of $\mathbb{F}_2[x_1, x_2, \ldots, x_m]$, in which the exponent of each variable x_i in each monomial equals 0 or 1. Write

$$\mathbb{F}_2^n = \{P_1, P_2, \dots, P_n\}.$$

Let ev denote the evaluation map from B_m to \mathbb{F}_2^n by the rule

$$ev(f) = (f(P_1), \dots, f(P_n)).$$

With this notation we define the Reed-Muller code of order r by

$$RM(r,m) = \{ev(f) \mid f \in B_m \& \deg(f) \le r\},\$$

where $\deg(f)$ is the global degree of the multivariate polynomial f (called the algebraic degree of the Boolean function that f represents).

We will use repeatedly the following Lemma, which is the case q = 2 of [11, Cor. 2].

Lemma 1. For all pairs of integers (r, m) with $0 \le r \le m$, the weight spectrum of RM(r+1, m+1) includes as a subset S + S, where S is the weight spectrum of RM(r, m), and $S + S = \{s_1 + s_2 \mid s_1, s_2 \in S\}$.

The proof of this lemma is particularly simple in our framework: we know (see [6]) that all codewords in RM(r+1, m+1) are obtained as the concatenation of a codeword u in RM(r+1, m) with a word of the form u+v, where v is a codeword in RM(r, m) (this is called the "(u,u+v) construction"). If u is itself taken in the subcode RM(r, m) of RM(r+1, m), then u and u+v are any codewords in this linear subcode and the result follows.

We shall refer to the following result as McEliece's congruence.

Theorem 1. The weights in RM(r,m) are multiples of $2^{\lfloor \frac{m-1}{r} \rfloor}$ [7]. This bound is tight in the sense that there is at least a codeword of RM(r,m) with weight $(2t+1)2^{\lfloor \frac{m-1}{r} \rfloor}$ for some integer t [1].

We will call the following deep result the Kasami-Tokura bound. It is a consequence of [6, Chapt. 15, th. 11], which also specifies explicitly A_w .

Theorem 2. Let w be a weight of RM(r,m) in the range $2^{m-r} \leq w < 2^{m-r+1}$. Let $\alpha = \min(r, m-r)$, and $\beta = \frac{m-r+2}{2}$. The weight w is of the form $w = 2^{m-r+1} - 2^{m-r+1-\mu}$, for μ in the range $1 \leq \mu \leq \max(\alpha, \beta)$. Conversely, for any such μ , there is a w of that form in the range $2^{m-r} \leq w < 2^{m-r+1}$.

We will also require the notion of BCH code of length n and designed distance d, hereby denoted by BCH(n, d). See [6, Chapt. 9] for a precise definition.

3 The weights of the Reed-Muller codes of length 2^m and order m-3

It is shown in [11, Theorem 16] by using Magma [8] and by induction on m that all the integers in $\{0, 2, 4, ..., 2^m\} \setminus \{2, 4, 6, 10, 2^m - 2, 2^m - 4, 2^m - 6, 2^m - 10\}$ are weights in RM(m-3, m) for $m \ge 6$.

The method used consists in applying Lemma 1, which says that the set of weights in RM(m-3,m) contains A + A where A is the set of weights in RM(m-4,m-1). The authors start then from RM(3,6), whose weights can be obtained by Magma [8]: $\{0, 2, 4, 6, 8, ..., 64\} \setminus \{2, 4, 6, 10, 54, 58, 60, 62\}$, and they proceed very simply by induction on m. We shall detail a little their proof (which is slightly informal) and show that the set above covers in fact all weights in RM(m-3,m).

Proposition 1. For every $m \ge 6$, the weights in RM(m-3,m) are the elements of $\{0, 2, 4, ..., 2^m\} \setminus \{2, 4, 6, 10, 2^m - 10, 2^m - 6, 2^m - 4, 2^m - 2\} = \{0, 8, 12 + 2i, 2^m - 8, 2^m\}$, where *i* ranges over consecutive integers from 0 to $2^{m-1} - 12$.

Proof. The result is correct for m = 6; assuming it is correct for $m \ge 6$, then denoting the set of weights by A, we have that A contains $\{0, 8, 12, 14, \ldots, 2^m - 14, 2^m - 12, 2^m - 8, 2^m\}$ and therefore A + A contains $\{0, 8, 12, 14, \ldots, 2^{m+1} - 14, 2^{m+1} - 12, 2^{m+1} - 8, 2^{m+1}\}$, since the case of $0, 8, 12, 14, \ldots, 2^m - 14, 2^m - 12, 2^m - 8, 2^m$ is covered by adding 0 to an element of A, the case of $2^m, 8 + 2^m$ is covered by adding 0 to an element of A, the case of $2^m, 8 + 2^m - 8, 2^m - 8, 2^m$ is covered by adding 0 to an element of A.

 2^m , $12 + 2^m$, $14 + 2^m$, ..., $2^{m+1} - 14$, $2^{m+1} - 12$, $2^{m+1} - 8$, 2^{m+1} is covered by adding 2^m to an element of A, and the remaining numbers $2^m - 10$, $2^m - 6$, $2^m - 4$, $2^m - 2$, $2^m + 2$, $2^m + 4$, $2^m + 6$, $2^m + 10$ are easily covered as well. By an induction using Lemma 1, all the numbers in $\{0, 8, 12 + 2i, 2^m - 12, 2^m - 8, 2^m\}$ are then weights in RM(m - 3, m). To complete the proof we show that these are the only possible weights in RM(m - 3, m).

The minimum distance of RM(m-3,m) being 8, and the code being stable under addition of the all-one vector (the constant Boolean function 1), the numbers 2, 4, 6, $2^m - 2$, $2^m - 4$, $2^m - 6$ cannot be weights, and according to Theorem 2 ([2, 4]), the only weights in the integral interval [8, 16) are the numbers of the form $16 - 2^i$ in this interval, and 10 is not among them. This completes the proof.

4 The weights of the Reed-Muller codes of length 2^m and order m - 4

Adapting the proof above to the order m - 4 needs to have the weights of RM(3,7) or better RM(4,8) (since we can observe that starting from RM(3,7) makes us lose some weights). It is impossible to run Magma [8] exhaustively in RM(3,7) (and a fortiori in RM(4,8)), because this code has size 2⁶⁴, but the weights of RM(4,8) are known from the Online Encyclopedia of Integer Sequences [12, http://oeis.org/A146976], and §6.1 below includes a direct determination of the weights of RM(3,7). The weights in RM(4,8)are

 $\{0, 16, 24, 28, 30, \dots, 226, 228, 232, 240, 256\},\$

where the dots represent all even integers between 32 and 224. This can be generalized as follows.

Proposition 2. For every $m \ge 8$, the set of all weights in RM(m-4,m) equals $\{0, 16, 24, 28 + 2i, 2^m - 24, 2^m - 16, 2^m\}$, where i ranges over the set of consecutive integers from 0 to $2^{m-1} - 28$.

Proof. The proof is similar to that of Proposition 1: using Lemma 1, we can show that the set of weights contains $\{0, 16, 24, 28 + 2i, 2^m - 28, 2^m - 24, 2^m - 16, 2^m\}$ and Theorem 2 tells us there is nothing else. Note that by Theorem 1 all weights are even.

Note that the weight distribution of RM(5,9) is given in Online Encyclopedia of Integer Sequences, see [12, http://oeis.org/A018897], and it confirms our result.

5 Numerical examples

In the present section, we show how some known weight spectra can be obtained as well as some new ones.

5.1 The weights of RM(3,7)

The weights of RM(2,6) are by the results in [6, Chapter 15] equal to:

 $S := \{0, 16, 24, 28, 32, 36, 40, 48, 64\}$

(actually, they are given in [12, OEIS A001726]).

A direct hand calculation or a computation in Magma [8] yields:

96, 100, 104, 112, 128}.

By McEliece's congruence (Theorem 1), the weights in RM(3,7) are multiples of 4. And by [4] the weight 20 and its complement to 64 are excluded. Hence, S + S equals the whole weight spectrum of RM(3,7).

5.2 The weights of RM(3,8)

The present subsection will show why the induction of Proposition 3 must start at m = 9. The weights in RM(2,7) are by the results in [6, Chapter 15] equal to

 $S := \{0, 32, 48, 56, 64, 72, 80, 96, 128\}$

(and this is confirmed by [12, OEIS A006006]). A direct hand calculation or a computation in Magma [8] yields

168, 176, 184, 192, 200, 208, 224, 256

By Theorem 2, this list is complete in the region $\{0, \ldots, 64\}$ and in its complement to 256. However all the elements in S + S are multiple of 8, while Theorem 1 states that the weights in RM(3,8) are multiples of 4, and some weights are not multiple of 8. Hence S + S is *strictly* included in the spectrum of RM(3,8), since McEliece's congruence is known to be tight [1]. This is confirmed by the known weight distribution [12, A146953] https://isec.ec.okayama-u.ac.jp/home/kusaka/wd/index.html which gives in addition the weights:

100, 108, 116, 124, 132, 140, 148, 156, 164,

that are congruent to 4 modulo 8.

These weights can be recovered by taking the intersection of RM(3,8) with the extension of BCH(255,19). The weight distribution of this intersection (a code of dimension 26, small enough to allow the use of the WeightDistribution command of Magma) is:

$$\begin{split} [<0,1>,<80,8>,<88,56>,<92,512>,<96,4939>,<100,30216>,\\ <104,159164>,<108,615184>,<112,1851060>,<116,4389152>,\\ <120,8126540>,<124,11733960>,<128,13287280>,<132,11733960>,\\ <136,8126540>,<140,4389152>,<144,1851060>,<148,615184>,\\ <152,159164>,<156,30216>,<160,4939>,<164,512>,<168,56>,\\ <176,8>,<256,1>]. \end{split}$$

Remark. We do not provide, properly speaking, a computer-free determination of the weight spectrum of RM(3,8). But we show how it can be derived in a reproducible way. \diamond

5.3 The weights of RM(4,9)

According to the previous subsection, the weight spectrum of RM(3,8) equals:

192, 200, 208, 224, 256.

A Magma calculation yields

480, 512.

By McEliece's congruence [7] the weights in RM(4,9) are multiples of 4. By Theorem 2 the weights 36, 40, 44, 52 and their complements to 512 are excluded, since

$$64 - 36 = 28, 64 - 40 = 24, 64 - 44 = 20, 64 - 52 = 12,$$

which are not powers of 2. However, still by Theorem 2 the integer $60 = 64 - 4 = 2^6 - 2^{6-4}$ is a weight of RM(4,9).

Note that this integer does not appear in the spectrum of RM(3,8) since then, in the notation of Theorem 2, $\mu \leq \lfloor \max(3, \frac{7}{2}) \rfloor = 3$, when $\mu = 4$ for the weight 60 of RM(4,9). This shows that the statement of Proposition 3 can only be valid for $m \geq 9$.

Hence $\{60\} \cup S + S$ is the whole spectrum of RM(4, 9).

5.4 The weights of RM(5, 10)

We summarize the partial knowledge we have of the spectrum of RM(5, 10).

Proposition 3. The weight spectrum of RM(5, 10) contains

- 1. the integer 62 and its complement to 1024, namely 962
- 2. all the even integers in the range [448, 576]

3. the set

 $\{0, 32, 48, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100, 104, 108, 112, 116, 120, 124, 128, 132, 136, 140, 144, 148, 152, 156, 160, 164, 168, 172, 176, 180, 184, 188, 192, 196, 200, 204, 208, 212, 216, 220, 224, 228, 232, 236, 240, 244, 248, 252, 256, 260, 264, 268, 272, 276, 280, 284, 288, 292, 296, 300, 304, 308, 312, 316, 320, 324, 328, 332, 336, 340, 344, 348, 352, 356, 360, 364, 368, 372, 376, 380, 384, 388, 392, 396, 400, 404, 408, 412, 416, 420, 424, 428, 432, 436, 440, 444, 448, 452, 456, 460, 464, 468, 472, 476, 480, 484, 488, 492, 496, 500, 504, 508, 512, 516, 520, 524, 528, 532, 536, 540, 544, 548, 552, 556, 560, 564, 568, 572, 576, 580, 584, 588, 592, 596, 600, 604, 608, 612, 616, 620, 624, 628, 632, 636, 640, 644, 648, 652, 656, 660, 664, 668, 672, 676, 680, 684, 688, 692, 696, 700, 704, 708, 712, 716, 720, 724, 728, 732, 736, 740, 744, 748, 752, 756, 760, 764, 768, 772, 776, 780, 784, 788, 792, 796, 800, 804, 808, 812, 816, 820, 824, 828, 832, 836, 840, 844, 848, 852, 856, 860, 864, 868, 872, 876, 880, 884, 888, 892, 896, 900, 904, 908, 912, 916, 920, 924, 928, 932, 936, 940, 944, 948, 952, 956, 960, 964, 968, 976, 992, 1024\}$

Proof.

- 1. By considering the special Boolean function $x_1x_2x_3x_4x_5 \oplus x_6x_7x_8x_9x_{10}$.
- 2. Follows by computing in Magma the weight distribution of the intersection between the extension of BCH(1023,157) and RM(5,10).
- 3. By Lemma 1 applied to the weight spectrum of RM(4,9) computed in the previous subsection.

6 Conclusion and open problems

In this note, we have derived the weight spectra of two infinite families of Reed-Muller codes. For further generalization, the base point of the recurrence might not be amenable to exact enumeration by computer and theoretical work will be necessary. The weight spectra of RM(4, 10) and RM(5, 10) are not known. (The weight spectra of RM(3, 9), which is given in http://oeis.org/A018895 and of RM(4, 10) and RM(5, 10) by the techniques in this note). This knowledge is necessary for starting an induction to determine the spectrum of RM(m-5, m).,

Based on the results of Proposition 1 and Proposition 2, it is natural to conjecture that the weight spectrum of RM(m-c, m) contains all the weights between the minimum distance 2^c and its complement to the length 2^m , that are authorized by McEliece's congruence and Kasami-Tokura's result. Since $\lfloor \frac{m-1}{m-c} \rfloor = 1$ for m > 2c - 1, we can formulate the following:

Conjecture: Let c be any positive integer. Then for m > 2c - 1, the weight spectrum of RM(m - c, m) is of the form:

$$\{0\} \cup A \cup B \cup C \cup \overline{B} \cup \overline{A} \cup \{2^m\},\$$

where:

- $A \subseteq [2^c, 2^{c+1}]$, is given by Kasami and Tokura [4],
- $B \subseteq [2^{c+1}, 2^{c+1} + 2^c]$, is given by Kasami, Tokura, and Azumi in [5, Page 392 and foll.],
- $C \subseteq [2^{c+1} + 2^c, 2^m 2^{c+1} 2^c]$, consists of consecutive even integers,
- \overline{A} stands for the complement to 2^m of A, and \overline{B} stands for the complement to 2^m of B.

This conjecture is verified for c = 1, 2, 3, 4. The first open case is c = 5.

In view of the ternary and quinary analogues of Theorem 15 of [11], namely Theorems 17 and 18 of [11], it is natural to ask for analogues of our results for Generalized Reed Muller codes of characteristics 3 and 5. However, an exact analogue of Theorem 2 in that context does not seem to be

available from the literature.

Acknowledgement. We thank the anonymous Reviewers for their useful comments; one of these provided the example of Boolean function mentioned in the proof of Proposition 3.

References

- Y.L. Borissov, On McEliece's result about divisibility of the weights in the binary Reed-Muller codes, Seventh International Workshop on Optimal Codes and Related Topics, September 6-12, 2013, Albena, Bulgaria pp. 47-52. http://www.moi.math.bas.bg/oc2013/a7.pdf
- [2] E. R. Berlekamp and N. J. A. Sloane. Restrictions on the weight distributions of the Reed-Muller codes. *Information and Control* 14, pp. 442-446, 1969.
- [3] C. Carlet. Boolean Functions for Cryptography and Coding Theory. Cambridge University Press, 2021.
- [4] T. Kasami and N. Tokura. On the weight structure of the Reed-Muller codes, *IEEE Transactions on Information Theory* 16, pp. 752-759, 1970.
- [5] T. Kasami, N. Tokura, and S. Azumi. On the Weight Enumeration of Weights Less than 2.5d of Reed-Muller Codes. *Information and Control*, 30:380–395, 1976.
- [6] F. J. MacWilliams and N. J. Sloane. The theory of error-correcting codes, North Holland. 1977.
- [7] R. J. McEliece. Weight congruence for p-ary cyclic codes. Discrete Mathematics, 3, pp. 177-192, 1972.
- [8] http://magma.maths.usyd.edu.au/calc/
- [9] S. Mesnager and A. Oblaukhov. Classification of the codewords of weights 16 and 18 of the Reed-Muller code RM(n-3,n). *IEEE Transactions on Information Theory* 68.2, pp. 940-952, 2021.

- [10] V. S. Pless, W. C. Huffman, Eds, R. A. Brualdi, assistant editor. Handbook of Coding Theory, Elsevier, 1998.
- [11] M. Shi, X. Li, A. Neri and P. Solé. How many weights can a cyclic code have? *IEEE Transactions on Information Theory*, vol. 66, no 3, p. 1449-1459, 2019.
- [12] N. J. Sloane. Online Encyclopedia of Integer Sequences (OEIS). http://oeis.org/wiki/Welcome