The weight spectrum of the Reed-Muller codes $RM(m-5, m)$ *

Claude Carlet[†]

Dept. of Informatics, University of Bergen, 5005 Bergen, Norway
Dept. of Mathematics, University of Paris 8, 93526 Saint-Denis, France

Abstract

The weight spectra (i.e. the lists of all possible weights) of the Reed-Muller codes $RM(r, m)$, of length 2^m and order r, are unknown for $r \in \{3, \ldots, m-5\}$ (and m large enough). Those of $RM(m-4, m)$ and $RM(m-3, m)$ have been determined very recently (but not the weight distributions, giving the number of codewords of each weight, which seem out of reach). We determine the weight spectrum of $RM(m-5, m)$ for every $m \geq 10$. We proceed by first determining the weights in $RM(5, 10)$. To do this, we construct functions whose weights are in the set $\{62, 74, 78, 82, 86, 90\}$, and functions whose weights are all the integers between 94 and $2^9 - 2 = 510$ that are congruent with 2 modulo 4 (those weights that are divisible by 4 are easier to determine and they are indeed known). This allows us to determine completely the weight spectrum, thanks to the well-known result due to Kasami, Tokura and Azumi, which precisely determines those codeword weights in Reed-Muller codes which lie between the minimum distance d and 2.5 times d, and thanks to the fact the weight spectrum is symmetric with respect to 2^9. Then we use this particular weight spectrum for determining that of $RM(m-5, m)$, by an induction on m.

This extended abstract is an excerpt of the full paper [3].

1 Introduction

Given $0 \leq r \leq m$, the Reed-Muller code $RM(r, m)$, of length 2^m and order r, is made of all m-variable Boolean functions f of algebraic degree at most r (or more precisely of the binary vectors of length 2^m that are the lists of values of $f(\mathbf{x})$ when $\mathbf{x} = (x_1, \ldots, x_m)$ ranges over $\mathbb{F}_2^m$ in some fixed order). All codeword weights in the Reed-Muller codes of length 2^m and orders $0, 1, 2, m-2, m-1, m$ are known (as well as the weight distributions of these codes). They are recalled for instance in [7] and in [4].
The low Hamming weights are also known in all Reed-Muller codes: Kasami and Tokura [5] have shown that, for $r \geq 2$, the only Hamming weights in $RM(r, m)$ occurring in the range $[2^{m-r}; 2^{m-r+1}[$ are of the form $2^{m-r+1} - 2^{m-r+1-i}$ where $i \leq \max(\min(m-r, r), \frac{m-r+2}{2})$.

Kasami, Tokura and Azumi determined later in [6] all the weights lying between the minimum distance $d = 2^{m-r}$ and 2.5 times d. The functions having such weights are characterized in this reference (all weights are described at pages 392 and following of the reference, and the corresponding functions are described under some conditions in its Table I).

The weight spectra (i.e., the sets of all possible codeword weights) of the codes $RM(r, m)$ are unknown for $3 \leq r \leq m-5$ (and therefore, their weight distributions are also unknown) but they have been recently determined in [4] for $r = m-4, m-3$, thanks to the fact that there is a simple way to determine many weights in $RM(r, m)$ from the weights in $RM(r-1, m-1)$; the weight spectra of $RM(m-c, m)$

---

*The full version of this work is to appear in IEEE Transactions on Information Theory. It has never been presented in a conference (only in a local seminar) .This research is supported by the Norwegian Research Council.

[†]Email: claude.carlet@gmail.com.

were then deduced for $c = 3, 4$, thanks to the Kasami-Tokura's results [5], which allowed to know that the numbers missing in the obtained lists could not be weights in these codes.

Reference [4] could not address the cases $c \geq 5$, mainly because the weights that are not divisible by 4 in $RM(5, 10)$ could not be determined. In the present paper, we solve the case $c = 5$, by constructing codewords in $RM(5, 10)$ achieving all the weights allowed by [6] and all those that are larger than $2.5d$ and smaller than 2^{m-1}, and thanks to an induction on m.

2 Preliminaries

The *Hamming weight* (in brief, the weight) of a binary vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$ is the size of its support $\{i \in \{1, \ldots, n\}; x_i \neq 0\}$. The *Hamming distance* between two vectors in $\mathbb{F}_2^n$ is the weight of their difference (that is, of their sum). Hence, since m-variable Boolean functions can be identified with binary vectors of length $n = 2^m$, the *Hamming weight* of an m-variable Boolean function f is the size of its support $\{\mathbf{x} \in \mathbb{F}_2^m; f(\mathbf{x}) \neq 0\}$, and the *Hamming distance* between two Boolean functions is the weight of their sum. A binary linear code of *length* n is an $\mathbb{F}_2$-subspace of $\mathbb{F}_2^n$. This allows to define its *dimension* (as an $\mathbb{F}_2$-vector space). Its *minimum distance* is the minimum Hamming distance between distinct codewords, that is (thanks to the linearity of the code) the minimum Hamming weight of the nonzero codewords.

The set of the codeword weights of a given linear code C will be called the *weight spectrum* of C, and for simplicity, we will sometimes write "the weights of C" instead of "the weights of the codewords in C" for the elements of its weight spectrum. The *weight distribution* of the code is the list of the numbers A_i, where A_i equals the number of codewords of weight i for $i \in \{0, \ldots, n\}$.

Given two integers m and $r \in \{0, \ldots, m\}$, the *Reed Muller* code $RM(r, m)$ of length $n = 2^m$ and order r is defined in terms of Boolean functions (see [7]): each m-variable Boolean function $f : \mathbb{F}_2^m \mapsto \mathbb{F}_2$ admits a unique representation as a polynomial in $\mathbb{F}_2[x_1, \ldots, x_m]/(x_1^2 + x_1, \ldots, x_m^2 + x_m)$, called the algebraic normal form (ANF) of f. We choose an order on $\mathbb{F}_2^m$, that is, we write $\mathbb{F}_2^m = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n\}$, and we denote by ev the evaluation map from the space of Boolean functions to $\mathbb{F}_2^n$ by the rule $ev(f) = (f(\mathbf{P}_1), \ldots, f(\mathbf{P}_n))$. Then $RM(r, m)$ equals $\{ev(f) \mid f \in B_m \text{ and } \deg(f) \leq r\}$, where B_m is the vector space of all m-variable Boolean functions and $\deg(f)$, called the algebraic degree of f, is the (global) degree of the ANF of f. Boolean function f has an odd Hamming weight if and only if it has (maximal) algebraic degree m.

The dimension of $RM(r, m)$ equals $\sum_{i=0}^{r} \binom{m}{i}$ and its minimum distance equals 2^{m-r}. The minimum weight codewords are the indicators of the $(m - r)$-dimensional affine subspaces of $\mathbb{F}_2^m$; up to affine equivalence, they equal $\prod_{i=1}^{r} x_i$ (two Boolean functions are called affine equivalent if one equals the composition of the other by an affine permutation).

The McEliece theorem gives a divisibility lower bound on the weights in $RM(r, m)$:

Theorem 1 (McEliece divisiblity theorem). *[8] The weights in $RM(r, m)$ are multiples of $2^{\lfloor \frac{m-1}{r} \rfloor}$.*

This bound is tight, as shown in [1]; more precisely, for each pair (r, m), there is at least one codeword of $RM(r, m)$ with weight equal to $2^{\lfloor \frac{m-1}{r} \rfloor}$ times an odd integer.

Another important result on Reed-Muller codes is the following (already evoked in the introduction):

Theorem 2 (Kasami-Tokura). *[5] Let w be a weight of some nonzero codeword in $RM(r, m)$ 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 $w = 2^{m-r+1} - 2^{m-r+1-i}$, for i in the range $1 \leq i \leq \max(\alpha, \beta)$. Conversely, for any such i, there is a w of that form in the range $2^{m-r} \leq w < 2^{m-r+1}$.*

This result has been extended in [6] into the characterization of all the weights of $RM(r, m)$ that are in the range $2^{m-r} \leq w < 2^{m-r+1} + 2^{m-r-1}$ (i.e. that lie between the minimum distance of the code and 2.5 times the minimum distance). It is impossible to summarize these results; we shall refer below to the pages in this reference where the results that we shall need can be found.

Notation: for every n, we denote respectively by $\mathbf{0}_n$ and $\mathbf{1}_n$ the all-0 and all-1 vectors of length n.

3 The weights of the Reed-Muller codes of length 2^m and order $m-5$

It is well-known that we obtain all the codewords in $RM(r,m)$ by concatenating any codeword u of $RM(r,m-1)$ and the sum of u and of a codeword v of $RM(r-1,m-1)$ (this is called the $(u,u+v)$ construction of $RM(r,m)$, see [7]). If we take u also in $RM(r-1,m-1)$, then u and $u+v$ range freely and independently in $RM(r-1,m-1)$. Hence, $RM(r,m)$ contains the concatenations of any two codewords of $RM(r-1,m-1)$ (which can also be seen directly by considering functions of the form $u(\mathbf{x}')+x_m v(\mathbf{x}')$, where u and v are two $(m-1)$-variable Boolean functions of algebraic degrees at most $r-1$ and $\mathbf{x}' \in \mathbb{F}_2^{m-1}$). This implies that the sums of two weights in $RM(r-1,m-1)$ are weights in $RM(r,m)$. This allowed in [4] to determine the weights of $RM(3,6)$ and $RM(4,8)$ and deduce by induction the weights of $RM(m-c,m)$ when $c \leq 4$.

But the weights in $RM(m-5,m)$ could not be determined. This would have needed to determine the weights in $RM(5,10)$. Indeed, determining the weights in the codes $RM(m-c,m)$ for a given $c>0$ needs in practice, for starting an induction, to determine the weights in the code $RM(m-c,m)$ for which m is the smallest such that $\left\lfloor \frac{m-1}{m-c} \right\rfloor$ (in the McEliece divisiblity theorem) has value 1, that is, $m=2c=2r$ (in which case the condition $i \leq \max(\min(m-r,r), \frac{m-r+2}{2})$ of Kasami-Tokura writes $i \leq c$). Taking m smaller than $2c$ allows by computing sums of two weights in $RM(m-c,m)$ to obtain only weights that are divisible by 4 in $RM(m+1-c,m+1)$. And only a half of the weights of $RM(5,10)$ could be determined in [4] (almost all weights that are not divisible by 4 missing).

For the reasons presented above, determining the weights in $RM(r,2r)$ that are divisible by 4 is easier than determining those which are not divisible by 4 (and divisible by 2): many of the former can be obtained by adding two weights from $RM(r-1,2r-1)$ if these weights are known, or from $RM(r-j,2r-j)$ where $j>1$ is the smallest value for which the weights are known. This is how they have been determined in [4] for $RM(5,10)$.
Let us then work on the most difficult part: the weights that are not divisible by 4.

3.1 The weights in $RM(5,10)$ that are congruent with 2 mod 4

Since using a computer for obtaining the weight spectrum of $RM(5,10)$ seems out of reach, we need to mathematically construct Boolean functions in 10 variables and of algebraic degree at most 5, whose Hamming weights can be determined and cover as many values allowed by [6] as possible (and are congruent with 2 mod 4). Of course, we only need to determine the weights up to $2^{m-1}-2$, since Reed-Muller codes being invariant by the complementation of their codewords to the all-one vector, their weight spectra are invariant by complement to 2^m.
We shall use the structure of the so-called Maiorana-McFarland functions (see e.g. [2]). Let m be a positive integer. An m-variable Boolean function is Maiorana-McFarland if there exist $2 \leq k \leq m$, $\phi : \mathbb{F}_2^{m-k} \mapsto \mathbb{F}_2^k$ and $g : \mathbb{F}_2^{m-k} \mapsto \mathbb{F}_2$ such that:

$$f(\mathbf{x},\mathbf{y}) = \mathbf{x}\cdot\phi(\mathbf{y}) + g(\mathbf{y}); \quad \mathbf{x} \in \mathbb{F}_2^k, \quad \mathbf{y} \in \mathbb{F}_2^{m-k},$$

where $(\mathbf{x},\mathbf{y})$ is the concatenation of the vectors $\mathbf{x}=(x_1,\dots,x_k)$ and $\mathbf{y}=(y_1,\dots,y_{m-k})$ and "$\cdot$" is an inner product in $\mathbb{F}_2^k$ (for instance the so-called usual inner product $\mathbf{x}\cdot\mathbf{x}' = x_1x_1' + \cdots + x_kx_k'$, where of course $\mathbf{x}'=(x_1',\dots,x_k')$). We assume $k \geq 2$ because for $k=1$, the corresponding Maiorana-McFarland functions are all m-variable Boolean functions, and the Maiorana-McFarland structure is then weak and does not help the study.

Such function f belongs to $RM(r,m)$ if and only if ϕ has algebraic degree at most $r-1$ (that is, all its coordinate functions have algebraic degree at most $r-1$) and g has algebraic degree at most r.

Considering the value $W_f(\mathbf{0}_k,\mathbf{0}_{m-k})$ of the Walsh transform W_f of function f (see e.g. [2]), we have:

$$2^m - 2w_H(f) = W_f(\mathbf{0}_k,\mathbf{0}_{m-k}) :=$$
$$\sum_{\mathbf{x}\in\mathbb{F}_2^k,\mathbf{y}\in\mathbb{F}_2^{m-k}} (-1)^{\mathbf{x}\cdot\phi(\mathbf{y})+g(\mathbf{y})} =$$
$$\sum_{\mathbf{y}\in\mathbb{F}_2^{m-k}} \left((-1)^{g(\mathbf{y})} \sum_{\mathbf{x}\in\mathbb{F}_2^k} (-1)^{\mathbf{x}\cdot\phi(\mathbf{y})}\right) = 2^k \sum_{\mathbf{y}\in\phi^{-1}(\mathbf{0}_k)} (-1)^{g(\mathbf{y})},$$

where $\phi^{-1}(\mathbf{0}_k)$ denotes the pre-image by ϕ of the zero vector in $\mathbb{F}_2^k$. Hence:

$$w_H(f) = 2^{m-1} - 2^{k-1} \sum_{\mathbf{y}\in\phi^{-1}(\mathbf{0}_k)} (-1)^{g(\mathbf{y})}. \tag{1}$$

We want this number to be congruent with 2 mod 4, which obliges to take $k=2$.

Let ϕ_1,ϕ_2 be the two coordinate functions of ϕ. We have $\phi^{-1}(\mathbf{0}_2) = \{\mathbf{y}\in\mathbb{F}_2^{m-2}; \phi_1(\mathbf{y}) = \phi_2(\mathbf{y}) = 0\}$. The indicator function of $\phi^{-1}(\mathbf{0}_2)$ equals then $(\phi_1(\mathbf{y})+1)(\phi_2(\mathbf{y})+1)$. According to what we recalled in Section 2, a Boolean function in $m-2$ variables has an odd Hamming weight if and only if it has (maximal) algebraic degree $m-2$. Hence, $\phi^{-1}(\mathbf{0}_2)$ has an odd size if and only if $\phi_1\phi_2+\phi_1+\phi_2$ has algebraic degree $m-2$.

We fix now $m=10$ and $r=5$ ($c=5$). The fact that $\phi_1\phi_2$ has algebraic degree $m-2=8$ implies that ϕ_1 and ϕ_2 both have algebraic degree 4 exactly.

We wish that $\phi^{-1}(\mathbf{0}_2)$ is as large as possible (then we can try to reach as many weights as possible with f by visiting as many Boolean functions g as possible). For this, we wish that the co-support of ϕ_1 (that is, the complement of its support) is as large as possible. We take then for ϕ_1 a minimum weight codeword in $RM(4,8)$. Up to affine equivalence, we can take $\phi_1(\mathbf{y}) = \prod_{j=1}^4 y_j$ (see [7, 2]). This ϕ_1 being chosen, we want that $\phi_1\phi_2$ has the algebraic degree 8 and that $\phi^{-1}(\mathbf{0}_2)$ has a maximum size. Let us then take $\phi_2(\mathbf{y}) = \prod_{j=5}^8 y_j$.

3.1.1 The weights achievable by f when $m=10$, $k=2$, $\phi_1(\mathbf{y}) = \prod_{j=1}^4 y_j$ and $\phi_2(\mathbf{y}) = \prod_{j=5}^8 y_j$

With such choices, we have:

$$\begin{aligned}\phi^{-1}(\mathbf{0}_2) &= \Big\{\mathbf{y}\in\mathbb{F}_2^8; \prod_{j=1}^4 y_j = \prod_{j=5}^8 y_j = 0\Big\}\\ &= (\mathbb{F}_2^4\setminus\{\mathbf{1}_4\})\times(\mathbb{F}_2^4\setminus\{\mathbf{1}_4\}).\end{aligned}$$

Then, according to (1), denoting by g' the restriction of g to $(\mathbb{F}_2^4\setminus\{\mathbf{1}_4\})^2$, by g_1 the restriction of g to $\{\mathbf{1}_4\}\times\mathbb{F}_2^4$ and by g_2 the restriction of g to $\mathbb{F}_2^4\times\{\mathbf{1}_4\}$, we have:

$$\begin{aligned} w_H(f) &= 2^9 - 2\sum_{\mathbf{y}\in(\mathbb{F}_2^4\setminus\{\mathbf{1}_4\})^2} (-1)^{g(\mathbf{y})}\\ &= 2^9 - 2\Big(15^2 - 2w_H(g')\Big)\\ &= 62 + 4w_H(g') \qquad (2)\\ &= 62 + 4w_H(g) - 4w_H(g_1) - 4w_H(g_2) + 4\,g(\mathbf{1}_8).\end{aligned}$$

The detailed explanations on how we obtained all the possible weights of g' when g belongs to $RM(5,8)$ can be found at URL:
`https://d197for5662m48.cloudfront.net/documents/publicationstatus/171039/preprint_pdf/5e3b1a34b6f649e6b532796b16033485.pdf`

The weights congruent with 2 mod 4 between 62 and 94 Considering the case where g has minimum nonzero weight 8 (i.e. g is the indicator of a 3-dimensional affine space A), and considering all possible cases, we have:

Lemma 3. *Let:*

$$f(\mathbf{x},\mathbf{y}) = x_1 \prod_{j=1}^{4} y_j + x_2 \prod_{j=5}^{8} y_j + g(\mathbf{y}); \quad \mathbf{x} \in \mathbb{F}_2^2, \quad \mathbf{y} \in \mathbb{F}_2^8,$$

where g is any minimum weight codeword in $RM(5,8)$. Then the set of weights of such codewords of $RM(5,10)$ includes $\{62, 74, 78, 82, 86, 90, 94\}$ and covers all the weights in $RM(5,10)$ that are congruent with 2 modulo 4 and between 62 and 94.

The weights congruent with 2 mod 4 between 96 and 126 Choosing now for g a codeword of $RM(5,8)$ having the three weights that come immediately after 8 when visiting the weight spectrum in ascending order, that is $16-4=12$, $16-2=14$ and 16 itself, we obtain:

Lemma 4. *Let f be defined as in Lemma 3, where g is the sum of two minimum weight codewords in $RM(5,8)$. Then the set of weights of such codewords of $RM(5,10)$ includes additionally to Lemma 3, the numbers: $98, 102, 106, 110, 114, 118, 122, 126$, and covers then all the weights in $RM(5,10)$ that are congruent with 2 modulo 4 and which lie between 98 and 126.*

The weights congruent with 2 mod 4 between 130 and 226 We now need to take a function g such that the weight w of g' is between 17 and 41. We have:

Lemma 5. *Let f be defined as in Lemma 3, where g is the sum of three to six minimum weight codewords in $RM(5,8)$ with disjoint supports. Then the set of weights of such codewords of $RM(5,10)$ includes additionally to Lemmas 3 and 4, all the numbers congruent with 2 modulo 4 and lying between 130 and 226.*

All remaining weights congruent with 2 mod 4

Lemma 6. *Let g be the 8-variable Maiorana-McFarland function:*

$$g(\mathbf{z},\mathbf{t}) = \mathbf{z} \cdot \psi(\mathbf{t}) + h(\mathbf{t}); \quad \mathbf{z},\mathbf{t} \in \mathbb{F}_2^4,$$

where ψ is any function from $\mathbb{F}_2^4$ to $\mathbb{F}_2^4$ and h is any Boolean function over $\mathbb{F}_2^4$. Let:

$$f(\mathbf{x},\mathbf{z},\mathbf{t}) = x_1 \prod_{j=1}^{4} (z_j+1) + x_2 \prod_{j=1}^{4} (t_j+1) + g(\mathbf{z},\mathbf{t});$$

$$\mathbf{x} \in \mathbb{F}_2^2, \quad \mathbf{z},\mathbf{t} \in \mathbb{F}_2^4.$$

Then the algebraic degree of any such 10-variable Boolean function f is at most 5 and the set of the weights of such functions includes all those integers between 230 and 510 that are congruent with 2 modulo 4.

3.2 The weight spectrum of $RM(5,10)$

Proposition 7. *The set of all weights in $RM(5,10)$ equals $\{0, 32, 48, 56, 60, 62, 64, 68, 72+2i, 2^{10}-68, 2^{10}-64, 2^{10}-62, 2^{10}-60, 2^{10}-56, 2^{10}-48, 2^{10}-32, 2^{10}\}$, where i ranges over the set of consecutive integers from 0 to 2^9-72.*

Proof. The result is deduced from Lemmas 3,4,5,6, the results of [5], and the facts that the spectrum is symmetric with respect to 512 and that, according to [4], all the numbers divisible by 4 between 56 and $2^{10}-56=968$ are weights in $RM(5,10)$. □

3.3 The weight spectrum of every code $RM(m-5,m)$ for $m \geq 10$

Theorem 8. *For every $m \geq 10$, the set of all weights in $RM(m-5,m)$ equals $\{0, 32, 48, 56, 60, 62, 64, 68, 72+2i, 2^m-68, 2^m-64, 2^m-62, 2^m-60, 2^m-56, 2^m-48, 2^m-32, 2^m\}$, where i ranges over the set of consecutive integers from 0 to $2^{m-1}-72$.*

The proof by an induction on $m \geq 10$ is omitted because of length limitation.

Open question: *Let c be any positive integer. For $m \geq 2c$, is the weight spectrum of $RM(m-c,m)$ 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 [5],*
- *$B \subseteq [2^{c+1}, 2^{c+1}+2^{c-1}]$, is given by Kasami, Tokura, and Azumi in [6, Page 392 and foll.],*
- *$C \subseteq [2^{c+1}+2^{c-1}, 2^m-2^{c+1}-2^{c-1}]$, consists of all 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.*

References

[1] 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] C. Carlet. *Boolean Functions for Cryptography and Coding Theory.* Cambridge University Press, 2021.

[3] C. Carlet. The weight spectrum of the Reed-Muller codes $RM(m-5,m)$. To appear in *IEEE Transactions on Information Theory.* 2024. 10.1109/TIT.2023.3343697

[4] C. Carlet and P. Solé. The weight spectrum of two families of Reed-Muller codes. *Discrete Mathematics* 346 (10), 113568, 2023. See also http://arxiv.org/abs/2301.13497.

[5] T. Kasami and N. Tokura. On the weight structure of the Reed-Muller codes, *IEEE Transactions on Information Theory* 16, pp. 752-759, 1970.

[6] 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.

[7] F. J. MacWilliams and N. J. Sloane. *The theory of error-correcting codes*, North Holland. 1977.

[8] R. J. McEliece. Weight congruence for p-ary cyclic codes. *Discrete Mathematics*, 3, pp. 177-192, 1972.

[9] List of weight distributions [from The On-Line Encyclopedia of Integer Sequences (OEIS)] https://oeis.org/wiki/List_of_weight_distributions