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

Claude Carlet

(1) University of Bergen, Norway

(2) LAGA, Universities of Paris 8 and Paris 13, CNRS, France

Outline

- Reed-Muller codes
- The weights in $RM(5, 10)$
- The weights in $RM(m-5, m)$ for every $m \geq 10$
- Extending the result to $RM(m-c, m)$ for $c \geq 6$?

Reed-Muller codes

Every m-variable Boolean function $f : \mathbb{F}_2^m \mapsto \mathbb{F}_2$ admits a unique representation as a polynomial:

$$\sum_{I \subseteq \{1,\ldots,m\}} a_I \prod_{i \in I} x_i \in \mathbb{F}_2[x_1, \ldots, x_m]/(x_1^2 + x_1, \ldots, x_m^2 + x_m);\ a_I \in \mathbb{F}_2$$

called the algebraic normal form (ANF) of f. The (global) degree of the ANF of f is called its algebraic degree.

The (binary) Reed-Muller code $RM(r, m)$, of length 2^m and order r, has for codewords the m-variable Boolean functions f of algebraic degree at most r (identified with binary vectors of length 2^m).

A linear code of *length* n over a field K is a K-subspace of K^n. This allows to define its *dimension* (as a K-vector space).

The *Hamming weight* (in brief, the weight) of a codeword is the size of its support. The *Hamming distance* between two codewords equals the Hamming weight of their difference.

The *minimum distance* of a code is the minimum Hamming distance between distinct codewords. When the code is linear, it equals the minimum Hamming weight of the nonzero codewords.

Weight spectrum of a code: set of all possible codeword weights.

Weight distribution: the number of codewords of each weight.

The Reed-Muller code $RM(r,m)$ is linear and has dimension $\sum\limits_{i=0}^{r}\binom{m}{i}$ and minimum distance 2^{m-r}.

The minimum weight codewords are the indicators of the $(m-r)$-dimensional affine subspaces of $\mathbb{F}_2^m$.

All weights in the Reed-Muller codes of length 2^m and orders $0,1,2,m-2,m-1,m$ are well-known (as well as the weight distributions of these codes).

The low Hamming weights are also known in all Reed-Muller codes: Kasami and Tokura have shown that, for $r \geq 2$, the only Hamming weights in $RM(r,m)$ 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 all the weights lying between the minimum distance $d = 2^{m-r}$ and 2.5 times d.

The McEliece theorem states that the weights in $RM(r,m)$ are multiples of $2^{\lfloor \frac{m-1}{r} \rfloor}$.

This divisibility bound is tight: for each pair (r,m), there is a codeword of $RM(r,m)$ with weight divisible by $2^{\lfloor \frac{m-1}{r} \rfloor}$ only.

The weight spectrum of $RM(r,m)$ is unknown for $3 \leq r \leq m-5$.

The weight spectrum of $RM(r,m)$ has been determined for $r \in \{m-4, m-3\}$ (C.C. and P. Solé, Discrete Mathematics, 2023):

- 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),
  - if we take u also in $RM(r-1,m-1)$, then u and $u+v$ range freely and independently in $RM(r-1,m-1)$,
  - $RM(r,m)$ contains then the concatenations of any two codewords of $RM(r-1,m-1)$, and
  - *sums of two weights in* $RM(r{-}1,m{-}1)$ *are weights in* $RM(r,m)$.
  This allows to obtain weights in $RM(m-c,m)$ by induction.

- The Kasami-Tokura's result allows to show that all weights in $RM(r, m)$ are obtained, when all even weights between $2d$ and $2^m - 2d$ are obtained.

Concretely:

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 result) has value 1.

That is, $m = 2c = 2r$. 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 $RM(m+1-c, m+1)$.

- Starting from the weights in $RM(3,6)$, obtained by Magma, and applying an induction, we have that, for every $m \geq 6$, the weights in $RM(m-3,m)$ are the elements of

$$\{0, 8, 12+2i, 2^m-8, 2^m\}; \quad i \in [0, 2^{m-1}-12].$$

- Starting from the weights in $RM(4,8)$, which are known from the Online Encyclopedia of Integer Sequences, and applying an induction as well, we have that for every $m \geq 8$, the set of all weights in $RM(m-4,m)$ equals:

$$\{0, 16, 24, 28+2i, 2^m-24, 2^m-16, 2^m\}; \quad i \in [0, 2^{m-1}-28].$$

The weight distributions of these codes seem out of reach currently (despite the fact that the weight distribution of the dual $RM(2,m)$ of $RM(m-3,m)$ is known: the number of codewords of weight 2^{m-1} in $RM(2,m)$ is too complex).

The weight spectra of $RM(m-c,m)$ could not be addressed for $c \geq 5$, mainly because the weights that are not divisible by 4 in $RM(5,10)$ could not be determined.

Here, we address the case $c=5$, by constructing codewords in $RM(5,10)$ achieving all the weights allowed by Kasami, Tokura and Azumi and by obtaining as weights all even numbers between $2.5\,d$ and $2^m-2.5\,d$, and thanks to an induction on m.

The weights in $RM(m-5,m)$ for every $m \geq 10$

The knowledge of the weight spectrum of $RM(m-4,m)$ does not help in determining that of $RM(m-5,m)$, even if the latter is a subset of the former.

Indeed, determining which of the weights in $RM(m-4,m)$ are also weights in $RM(m-5,m)$ is precisely what is difficult, because all even numbers between 28 and 2^m-28 are weights in $RM(m-4,m)$.

As we explained, determining the weights in $RM(c,2c)$ that are divisible by 4 is easier than determining those which are not divisible by 4 (and divisible by 2).

The weights in $RM(5,10)$

Of course, we only need to determine the weights up to $2^{m-1}-2$.

The Maiorana-McFarland class is made of the functions

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

where $2 \le k \le m$, $(\mathbf{x},\mathbf{y})$ is the concatenation of $\mathbf{x} = (x_1,\ldots,x_k)$ and $\mathbf{y} = (y_1,\ldots,y_{m-k})$, $\phi : \mathbb{F}_2^{m-k} \mapsto \mathbb{F}_2^k$ and $g : \mathbb{F}_2^{m-k} \mapsto \mathbb{F}_2$, and where "$\cdot$" is an inner product in $\mathbb{F}_2^k$ (for instance the usual inner product $\mathbf{x}\cdot\mathbf{x}' = x_1x_1' + \cdots + x_kx_k'$).

f belongs to $RM(r,m)$ if and only if ϕ has algebraic degree at most $r-1$ and g has algebraic degree at most r.

We have:

$$2^m - 2w_H(f) = W_f(\mathbf{0}_k, \mathbf{0}_{m-k}) :=$$

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

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

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

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

We fix now $m = 10$ and $r = 5$.

We wish that $\phi^{-1}(\mathbf{0}_2)$ is as large as possible for being able to reach as many weights as possible through proper choices of g.

We take then $\phi_1(\mathbf{y}) = \prod_{j=1}^{4} y_j$ and $\phi_2(\mathbf{y}) = \prod_{j=5}^{8} y_j$.

With such choices, we have:

$$\phi^{-1}(\mathbf{0}_2) = (\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\}) \times (\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\}) = (\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\})^2.$$

Denoting by g' the restriction of g to $(\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\})^2$, we have:

$$w_H(f) = 62 + 4\, w_H(g').$$

- *When* $g \in RM(5,8)$ *has (minimum) weight 8 (i.e. is the indicator of a 3-dimensional affine space* A *in* $\mathbb{F}_2^8$*)*:

Case (i): $A \subset (\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\})^2$, e.g. $A = \langle \mathbf{e}_1 + \mathbf{e}_5, \mathbf{e}_2 + \mathbf{e}_6, \mathbf{e}_3 + \mathbf{e}_7 \rangle \rightarrow$ weight $62 + 4 \cdot 8 = 94$.

Case (ii): $A \cap \Big((\mathbb{F}_2^4 \times \{\mathbf{1}_4\}) \cup (\{\mathbf{1}_4\} \times \mathbb{F}_2^4) \Big)$ is an affine subspace of $\mathbb{F}_2^4 \times \{\mathbf{1}_4\})$ or of $\{\mathbf{1}_4\} \times \mathbb{F}_2^4 \to$ weights $62 + 4 \cdot 7 = 90$, $62 + 4 \cdot 6 = 86$ and $62 + 4 \cdot 4 = 78$.

Case (iii): $A \cap \Big((\mathbb{F}_2^4 \times \{\mathbf{1}_4\}) \cup (\{\mathbf{1}_4\} \times \mathbb{F}_2^4) \Big)$ is the union of two affine spaces, one included in $\mathbb{F}_2^4 \times \{\mathbf{1}_4\}$ and one included in $\{\mathbf{1}_4\} \times \mathbb{F}_2^4$.

Case (iii).1: These two affine spaces have the point $\mathbf{1}_8$ in common $\to$ weights $62 + 4 \cdot 5 = 82$ and $62 + 4 \cdot 3 = 74$.

Case (iii).2: These two affine spaces are disjoint $\to$ no new weight

$\Rightarrow$ weights $\{62, 74, 78, 82, 86, 90, 94\}$.

The weights 66 and 70 are missing as expected (not in the list of Kasami-Tokura-Azumi).

In particular, we reach all those weights in $RM(5,10)$ allowed by Kasami-Tokura-Azumi that are not divisible by 4.

- *When* g *is a codeword of* $RM(5,8)$ *having one of the three weights that come after 8 :* $16-4=12$, $16-2=14$ *and 16, i.e.* g *is the sum of indicators of two 3-dimensional affine spaces; the intersection of these two spaces can be*:

- a 1-dimensional affine space, whose intersection with $\mathbb{F}_2^8 \setminus (\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\})^2 = (\mathbb{F}_2^4 \times \{\mathbf{1}_4\}) \cup (\{\mathbf{1}_4\} \times \mathbb{F}_2^4)$ has at most 2 elements,

- or a singleton, whose intersection with $(\mathbb{F}_2^4 \times \{\mathbf{1}_4\}) \cup (\{\mathbf{1}_4\} \times \mathbb{F}_2^4)$ has at most 1 element,

- or the empty set.

This provides the following additional values for the weight of g': $\{9, 10, 11, 12, 13, 14, 15, 16\}$ and by multiplying by 4 and adding 62, we obtain the weights: $98, 102, 106, 110, 114, 118, 122, 126$. This covers then all the weights in $RM(5, 10)$ that are congruent with 2 modulo 4 and which lie between 98 and 126.

- *When g is the sum of (at most) six functions $g_1, g_2, g_3, \ldots$ such that the corresponding functions $g'_1, g'_2, g'_3, \ldots$ have disjoint supports and their weights $w_1, w_2, w_3 \ldots$ sum to w*

This provides as weights all the numbers congruent with 2 modulo 4 and lying between 130 and 226.

- Obtaining all remaining weights congruent with 2 mod 4

Thanks to to a translation by $\mathbf{1}_8$, we change $(\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\}) \times (\mathbb{F}_2^4 \setminus \{\mathbf{1}_4\})$ into $(\mathbb{F}_2^4 \setminus \{\mathbf{0}_4\}) \times (\mathbb{F}_2^4 \setminus \{\mathbf{0}_4\})$.

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 $\mathbb{F}_2^4$. And we still take:

$$f(\mathbf{x}, \mathbf{z}, \mathbf{t}) = x_1 \prod_{j=1}^{4} (z_j+1) + x_2 \prod_{j=1}^{4} (t_j+1) + g(\mathbf{z}, \mathbf{t}); \quad \mathbf{x} \in \mathbb{F}_2^2, \mathbf{z}, \mathbf{t} \in \mathbb{F}_2^4.$$

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 include all those integers between 230 and 510 that are congruent with 2 modulo 4.

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

Proposition 1. *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 $i \in [0, 2^9 - 72]$.

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

Theorem 2. *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 $i \in [0, 2^{m-1}-72]$.

Proof by induction.

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,
- $B \subseteq [2^{c+1}, 2^{c+1} + 2^{c-1}]$, is given by Kasami, Tokura, and Azumi,
- $C \subseteq [2^{c+1} + 2^{c-1}, 2^m - 2^{c+1} - 2^{c-1}]$, 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 ?

Extending the result to $RM(m-c,m)$ for $c \geq 6$?

The smallest value of m we can take, which may allow determining all weights in $RM(m-c,m)$, is $m=2c$.
We can again consider functions

$$f(\mathbf{x},\mathbf{y}) = \mathbf{x}\cdot\phi(\mathbf{y}) + g(\mathbf{y});\ \mathbf{x}\in\mathbb{F}_2^k,\ \mathbf{y}\in\mathbb{F}_2^{m-k}, \tag{1}$$

with $k=2$, $m-k=2(c-1)$, where ϕ equals (ϕ_1,ϕ_2), with $\phi_1(\mathbf{y}) = \prod_{i=1}^{c-1} y_i$ and $\phi_2(\mathbf{y}) = \prod_{i=c}^{2c-2} y_i$, and $g\in RM(c,2c-2)$.

We have $w_H(f) = 2^{c+1}-2+4w_H(g')$ where g' is the restriction of g to $(\mathbb{F}_2^{c-1}\setminus\{\mathbf{1}_{c-1}\})^2$.

We could check in the case of $RM(6,12)$ that the weights of g' that can be reached this way are 7,8,9,11,12,13,14,15,16.

The number 10 is then missing.

Hence, the weight $2^7 - 2 + 4 \cdot 10 = 166$ is missing for f.

We could reach it with f equal to the sum of three minimum weight codewords of $RM(6,12)$.

We continue with

$$f(\mathbf{x},\mathbf{z},\mathbf{t}) = x_1\prod_{j=1}^{c-1}(z_j+1)+x_2\prod_{j=1}^{c-1}(t_j+1)+g(\mathbf{z},\mathbf{t});\ \mathbf{x}\in\mathbb{F}_2^2,\ \mathbf{z},\mathbf{t}\in\mathbb{F}_2^{c-1}$$

$$g(\mathbf{z},\mathbf{t}) = \mathbf{z}\cdot\psi(\mathbf{t})+h(\mathbf{t});\ \mathbf{z},\mathbf{t}\in\mathbb{F}_2^{c-1},$$

where ψ is any function from $\mathbb{F}_2^{c-1}$ to $\mathbb{F}_2^{c-1}$ and h is any Boolean function over $\mathbb{F}_2^{c-1}$.

Taking $\psi(\mathbf{0}_{c-1}) = \mathbf{0}_{c-1}$ and $h(\mathbf{0}_{c-1}) = 0$, we have:

$$w_H(f) = 2^{c+1} - 2 + 2^{2c-1} - 2^c \sum_{\mathbf{t} \in \psi^{-1}(\mathbf{0}_{c-1})} (-1)^{h(\mathbf{t})} - 4w_H(h).$$

Work in progress

Conclusion

Determining the weight distributions of the Reed-Muller codes $RM(r, m)$ for $r \in \{3, \ldots, m-4\}$ and every m seems out of reach.

The weight distribution of $RM(m-3, m)$ may be found if we ever manage to get a simpler expression for the number of codewords of weight 2^{m-1} in $RM(2, m)$, but there has been no progress on this for half a century.

The weight spectra of $RM(r, m)$ (without the knowledge of the number of codewords of each weight) are currently what seems the least unattainable for $r \in \{3, \ldots, m-4)$.

Determining the weight spectra of $RM(m-3,m)$ for $m \geq 6$ and $RM(m-4,m)$ for $m \geq 8$ was not too difficult.

Determining the weight spectrum of $RM(m-5,m)$ for $m \geq 10$ has needed more work, but it could be done thanks to the fact that, surprisingly, the Maiorana-McFarland construction (with rather specific parameters) allows to reach all weights.

This is not the case for $RM(m-c,m)$ with $c \geq 6$ and $m \geq 2c$, whose weight spectra may pose big problems.

If the weight spectra of all codes $RM(m-c,m)$ for $c \geq 6$ and $m \geq 2c$ are determined (which would be a breakthrough), this will give the weight spectra of all $RM(r,m)$ with $r \geq \frac{m}{2}$.

Then determining those such that $r < \frac{m}{2}$ will remain an open question. In particular, the weight spectrum of $RM(3,m)$ will remain wide open, probably.

An old result (1990) shows that the weights of all Boolean functions are simply related to the weights of cubic Boolean functions (in many more variables); this shows that the weights of cubics are complex, contrary to those of quadratic functions (even if, according to McEliece's theorem, those weights are all divisible by $2^{\lfloor \frac{m-1}{3} \rfloor}$).