Weight Distribution of the Binary Reed-Muller Code $\mathcal{R}(4,9)$

Miroslav Markov

Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, Sofia, Bulgaria

joint work with Yuri Borissov

WCC2024: The XIII-th International Workshop on Coding and Cryptography

Perugia, Italy, June 17-21, 2024

19.06.2024

Outline

Some Definitions and Notations (1)

- **weight distribution**

Definition 1.

The weight distribution of a code $\mathbf{C}$ of length n is the vector $W(\mathbf{C}) = (W_0, \ldots, W_n)$, where W_i denotes the number of codewords of Hamming weight i.

- **weight spectrum**

Definition 2.

The weight spectrum of a code $\mathbf{C}$ with weight distribution $W(\mathbf{C}) = (W_0, \ldots, W_n)$ is the set $\{i : 0 \leq i \leq n, W_i > 0\}$.

- the simplest form of **weight enumerator**

Definition 3.

The following polynomial in the indeterminate z: $\mathcal{W}[z; \mathbf{C}] = \sum_{i=0}^{n} W_i z^i$ is called a weight enumerator of the code $\mathbf{C}$ with weight distribution $W(\mathbf{C}) = (W_0, \ldots, W_n)$.

Some Definitions and Notations (2)

We assume familiarity with notions of:

- **Boolean function**, **Algebraic Normal Form**, the **General Affine group** $GA(m)$ and its subgroup the **General Linear group** $GL(m, 2)$ acting on $\mathbb{F}_2^m$;
- The set of all Boolean functions in m variables, will be denoted by $\mathcal{B}_m$.

- **binary Reed-Muller code**

Definition 4.

For $0 \leq r \leq m$, the $r-$th order binary Reed-Muller (or RM) code $\mathcal{R}(r, m)$ is the set of all vectors $\mathbf{f}$ of length $n = 2^m$ whose corresponding $f \in \mathcal{B}_m$ are of algebraic degree at most r.

- Recall:

Statement 5.

For any m and any $r, 0 \leq r \leq m$, the binary RM code $\mathcal{R}(r, m)$ is a linear $[n, k, d]$ code with:

- *length $n = 2^m$, dimension $k = \sum_{i=0}^{r} \binom{m}{i}$ and minimum distance $d = 2^{m-r}$;*
- *the dual of $\mathcal{R}(r, m)$ is $\mathcal{R}(m - r - 1, m)$;*
  in particular, for any $s \geq 1$ the code $\mathcal{R}(s, 2s + 1)$ is a self-dual code.

- The action of $A \in GA(m)$ on a Boolean function $f(\mathbf{x})$ is denoted by $f \circ A$, i.e:

$$f \circ A(\mathbf{x}) = f(A(\mathbf{x})).$$

- Recall:

Definition 6.

The cosets $C_1 = f_1 + \mathcal{R}(r,m)$ and $C_2 = f_2 + \mathcal{R}(r,m)$ of $\mathcal{R}(r,m)$ with $f_1, f_2 \in \mathcal{B}_m$ are called affine equivalent if there exists a transformation $A \in GA(m)$: $f_2 = f_1 \circ A$.

- The following well-known fact is extensively used in our work (see, e.g., [4]):

Statement 7.

The weight enumerators of two affine equivalent cosets of a Reed-Muller code are identical.

Motivation

- By [13] one concludes that for $m \leq 9$ the **only so far unknown** is the (exact) weight distribution of $\mathcal{R}(4,9)$:
  - $\mathcal{R}(4,9)$ was listed among the smallest Reed-Muller codes whose weight distributions were unknown (in 1977) [11, p. 447];
  - The weight spectrum of that code has been found in [1];
  - To our knowledge there have been very few attempts to find (exact) weight distribution of $\mathcal{R}(4,9)]$, namely:
    - Since $\mathcal{R}(4,9)$ is a doubly even binary self-dual code, the general form of weight enumerators of such codes is known from A. M. Gleason's work (see, e.g., [11, Ch.19]) might be of help. But, although this approach has been successful for shorter RM codes requiring modest efforts for computing, its application to the code of interest needs more intrinsic knowledge than presented in [6, 7] (see, [2, Ch. 11] for details).
    - D. V. Sarwate has evaluated that the methods from [12] are not applicable to $\mathcal{R}(4,9)$ since there are too many equivalence classes of cosets of the desired kind to be useful;

For $0 \leq r \leq m$, denote by $\mathcal{H}^{(r)}(m)$ the set of all homogeneous polynomials on m binary variables of algebraic degree r adjoined with the 0.

Theorem 8.

([12, Sarwate 5.12]) For $0 \leq r \leq m$, it holds:

$$\mathcal{W}[z; \mathcal{R}(r+2, m+2)] = \sum_{p \in \mathcal{H}^{(r+2)}(m+1)} \mathcal{W}^2[z; p + \mathcal{R}(r+1, m+1)].$$

Theorem 9.

([12, Sarwate 5.13])
Let $p = e + f x_{m+1}$, with given $e \in \mathcal{H}^{(r+2)}(m)$ and $f \in \mathcal{H}^{(r+1)}(m)$. Then the weight enumerator of the coset $\mathcal{C}(p) = p + \mathcal{R}(r+1, m+1)$ equals to:

$$\sum_{g \in \mathcal{H}^{(r+1)}(m)} \mathcal{W}[z; e + g + \mathcal{R}(r, m)] \cdot \mathcal{W}[z; e + f + g + \mathcal{R}(r, m)].$$

- The affine equivalence classification of the cosets of RM codes is useful in studying important coding theoretical and cryptographic properties of Boolean functions, e.g., the covering radii. Recently, the interest in that topic has been renewed by [3] which provides (among other things) a method to classify $\mathcal{B}_7$;
- In our work, we make use of:
  - Langevin & Leander's classification [10] of the quotient space $\mathcal{R}(4,8)/\mathcal{R}(3,8)$ under the action of $GL(8,2)$, i.e., the classification of the Boolean quartic forms in eight variables;
  - Gillot & Langevin's classification [3] of the cosets of $\mathcal{R}(2,7)$ in $\mathcal{R}(4,7)$.

The refined approach: rationalle (1)

Let $n(k,m)$ be the number of linear equivalence classes of the quotient space $\mathcal{R}^*(k,m) = \mathcal{R}(k,m)/\mathcal{R}(k-1,m)$, i.e. the number of orbits to which $\mathcal{R}^*(k,m)$ is partitioned under the action of $GL(m,2)$. Assume that some numbering of these classes is fixed.

Corollary 10.

Let $p_i \in \mathcal{H}^{(r+2)}(m+1)$ and L_i be a representative and size, respectively, of the $i-$th linear equivalence class in $\mathcal{R}^*(r+2,m+1)$. Then, it holds:

$$\mathcal{W}[z;\mathcal{R}(r+2,m+2)] = \sum_{i=1}^{n(r+2,m+1)} L_i \mathcal{W}^2[z;p_i+\mathcal{R}(r+1,m+1)]. \quad (1)$$

Proof.

The claim is an immediate consequence of Theorem 8 and Statement 7. □

The refined approach: rationalle (2)

Corollary 11.

For given $e \in \mathcal{H}^{(r+2)}(m)$*, let* $\mathcal{H}^{(r+1)}(m)$ *is partitioned into blocks* $G_i, 1 \leq i \leq s$*, such that if* $g \in G_i$ *the enumerator* $\mathcal{W}[z; e+g+\mathcal{R}(r,m)]$ *is a (distinct) constant polynomial* $w_i(z)$*. Then the weight enumerator of the coset* $\mathcal{C}(p) = p + \mathcal{R}(r+1, m+1)$ *where* $p = e + f x_{m+1}$ *with* $f \in \mathcal{H}^{(r+1)}(m)$*, can be expressed by*

$$\sum_{i=1}^{s} w_i(z) \left(\sum_{g \in G_i} \mathcal{W}[z; e + f + g + \mathcal{R}(r,m)] \right). \tag{2}$$

Proof.

Follows by Theorem 9 rearranging the summands and putting outside brackets the common multipliers $w_i(z), 1 \leq i \leq s$. □

The refined approach: rationalle (3)

Corollaries 10-11 make feasible the computation of $\mathcal{W}[z;\mathcal{R}(4,9)]$, namely:

- Corollary 10 reduces the number of needed weight enumerator computations from the straightforward $|\mathcal{H}^{(4)}(8)| = 2^{\binom{8}{4}} = 2^{70}$ to the **reasonable** $n(4,8) = 999$.
- The affine equivalence classification of $\mathcal{R}(4,7)/\mathcal{R}(2,7)$ allows to substantiate the usage of Corollary 11 (thoroughly explained in general settings on the next slide).

- Recall:

Definition 12.

The subgroup $\mathcal{S}t(e)$ of $GA(m)$ that fixes given $e \in \mathcal{H}^{(r+2)}(m)$ is called stabilizer of e in $GA(m)$, i.e., for each $A \in \mathcal{S}t(e)$ it holds $e \circ A \in e + \mathcal{R}(r+1, m)$.

- For given $e \in \mathcal{H}^{(r+2)}(m)$, consider the partition $\Delta(e)$ of the cosets of form $e + g + \mathcal{R}(r, m), g \in \mathcal{H}^{(r+1)}(m)$ under the action of the stabilizer $\mathcal{S}t(e)$. By Statement 7, we can talk for "orbit weight" enumerator: the common weight enumerator of all orbit members. Moreover, we can constitute efficiently the coarse partition $\Delta'(e) = \{G_i, 1 \leq i \leq s\}$ from Corollary 11, by merging the orbits with identical weight enumerators (computed in advance on chosen orbit representatives of $\Delta(e)$).
- So, the number of needed polynomial multiplications to compute expr. (2) is reduced to the number of distinct orbit enumerators while that of polynomial additions is, of course, retained to (almost) $2^{\binom{m}{r+1}}$.

- Let $\mathcal{E}(4,7)$ be the set of representatives of the 12 linear equivalence classes of $\mathcal{R}^*(4,7)$ given in [9].

  For each $e \in \mathcal{E}(4,7)$, we perform in advance the following three tasks:

  - $\mathcal{T}1$: Constitute and store the orbits of the partition $\Delta(e)$ ("orbit algorithm" [5]);
  - $\mathcal{T}2$: Compute the weight enumerators of the cosets $e + g + \mathcal{R}(2,7)$ when g runs over a set of representatives of $\Delta(e)$'s orbits (by exhaustive generation of $\mathcal{R}(2,7)$ based on some Gray code);
  - $\mathcal{T}3$: Merge the orbits with identical weight enumerators to get the coarse $\Delta'(e)$.

- **Note**: Data arrangement enables for given $f \in \mathcal{H}^{(3)}(7)$ to look up the identifier of the orbit (block) in $\Delta(e)$ ($\Delta'(e)$) containing f.

The refined approach: pre-computing (2)

Table: Sizes of partitions $\Delta(e)$ and $\Delta'(e)$

$e \in \mathcal{E}(4,7)$: ANF's according to ([9])	$\lvert\Delta(e)\rvert$	$\lvert\Delta'(e)\rvert$
0	12	12
4567	63	52
1235+1345+1356+1456+2346+2356+2456	130	112
2367+4567	289	182
1237+4567	480	306
1257+1367+4567	730	395
1237+1247+1357+2367+4567	204	157
1236+1257+1345+1467+2347+2456+3567	1098	675
1236+1356+1567+2357+2467+2567+3456	1340	811
1367+2345+2356+3456+4567	6449	2170
1234+1237+1267+1567+2345+3456+4567	23988	3377
1236+1367+1567+2345+3456+3457+3467	33660	4636

- We have developed and implemented two algorithms (see, the *Proceedings*):
  - **Algorithm 1** which returns $\mathcal{W}[z; p + \mathcal{R}(3,8)]$ where $p = e + fx_8$ for given inputs $e \in \mathcal{E}(4,7)$ and $f \in \mathcal{H}^{(3)}(7)$ (using expr. (2) in Corollary 11);
  - **Algorithm 2** which computes the sum in Corollary 10, and thus $\mathcal{W}[z; \mathcal{R}(4,9)]$.
  - **Note**:

    The second algorithm requires a list $\mathcal{S}$ of pairs: (representative p_i, class size L_i) for the $i-$th class of the classification of $\mathcal{R}^*(4,8)$ where $p_i = e + f_i x_8$ for some $e \in \mathcal{E}(4,7)$ and $f_i \in \mathcal{H}^{(3)}(7), 1 \leq i \leq 999$.

- To provide a list $\mathcal{S}$, we make use of data present in [8]. However, there p'_i are of the form $e' + f'_i x_8$ where e's constitute different set $\mathcal{E}'(4,7)$ of representatives of the 12 classes of $\mathcal{R}^*(4,7)$;
- To adjust, we follow a procedure (derived by [4]) consisting of 3 steps:
  - Form the sets $\mathcal{E}'(3,7), \mathcal{E}(3,7)$ of duals of the forms in $\mathcal{E}'(4,7), \mathcal{E}(4,7)$, respectively;
  - Match the linearly equivalent pairs $(\overline{e}', \overline{e}) \in \mathcal{E}'(3,7) \times \mathcal{E}(3,7)$ using the invariants given in [4, pp. 115-117]), so the pairs in the original sets are matched, too;
  - For each matched pair $(e', e) \in \mathcal{E}'(4,7) \times \mathcal{E}(4,7)$, generate at random a nonsingular (7×7) matrix $\mathbf{A}$ and check the condition $e' \circ \mathbf{A} \in e + \mathcal{R}(3,7)$ until such matrix is obtained.
- The last step is carried out efficiently due to relatively large stabilizers sizes, e.g., the smallest is of size $9216 \approx 2^{13.17}$ while $|GL(7,2)| \approx 2^{47.21}$;
- Finally, acting on $f'_i, 1 \leq i \leq 999$, by the obtained linear transitions, we get a needed list $\mathcal{S}$.

The refined approach: pre-computing (4)

Table: The matching between $\mathcal{E}'(3,7)$ and $\mathcal{E}(3,7)$

$\mathcal{E}'(3,7)$	$\mathcal{E}(3,7)$
0	0
123	123
127+136+145	137+147+157+237+247+267+467
125+134	123+145
126+345	123+456
126+135+234	123+245+346
135+146+235+236+245	123+145+246+356+456
127+136+145+234	124+137+156+235+267+346+457
125+134+135+167+247+357	127+134+135+146+234+247+457
123+247+356	123+127+147+167+245
147+156+237+246+345	123+127+167+234+345+456+567
127+146+236+345	125+126+127+167+234+245+457

The refined approach: evaluating computational costs of the pre-computing

Briefly:

- the computational cost of task $\mathcal{T}1$ is $|\mathcal{H}^{(3)}(7)| \times \sum_{e \in \mathcal{E}(4,7)} |Sg(e)| = 2^{35} \times 26 \approx 2^{40}$ affine transformations where $Sg(e)$ denotes the set of generators of the stabilizer $\mathcal{S}t(e)$;
- the computational cost of task $\mathcal{T}2$ is in total proportional to the product $68443 \times 2^{29} \approx 2^{45}$ with the first factor being the number of classes of $\mathcal{R}(4,7)/\mathcal{R}(2,7)$ and the second being the size of $\mathcal{R}(2,7)$;
- the compressed storing of orbits and data arrangement into RAM needs at most 124 GB of memory.

The set of linear equivalence classes of $\mathcal{R}^*(4,8)$ is naturally partitioned into subsets of cardinalities $\mu(e)$ for fixed $e \in \mathcal{E}(4,7)$ and distinct $f \in \mathcal{H}^{(3)}(7)$ (see, [8]):

$$\overline{\mu} = (3,2,21,15,89,56,10,7,502,1,1,292)$$

- By Corollaries 10-11, one can easily deduce the following estimates, i.e.:
  - necessary multiplications of degree 128 polynomials:

$$\sum_{e \in \mathcal{E}(4,7)} \mu(e) \times |\Delta'(e)| = 1827252 \approx 2^{21};$$

  - necessary additions of degree 128 polynomials:

$$n(4,8) \times 2^{\binom{7}{3}} = 999 \times 2^{35} \approx 2^{45};$$

  - 999 squarings of degree 256 polynomials; and some additional operations of negligible cost, of course.

Conclusion

- Recent advances in the classification of Boolean functions [3],[10] and the utilization of modern high performance computers make feasible the application of Sarwate's approach [12] to determining **exact** weight distribution of $\mathcal{R}(4,9)$;
- However, we should admit that it may not be doable to push this line of research much further due to the enormous increase in computational burden with code length.

Weight Distribution of $\mathcal{R}(4,9)$

0	512	1
32	480	52955952
48	464	919315326720
56	456	271767121346560
60	452	860689275027456
64	448	89163020044002040
68	444	1777323352931696640
72	440	64959328938397057024
76	436	2094952122987829002240
80	432	86129855718211879936768
84	428	371838722874329360498 6880
88	424	21640767440064774686146 5600
92	420	1595894539503502293205411 4304
96	416	157096476311405305549517438 9136
100	412	2077552444573037520356371542 83520
104	408	34164336816436357675455725024 378880
108	404	5992987676360073735151889707696 128000
112	400	98321792181003426335755247508902 1004288
116	396	1408811591686009227109831306254561 63782656
120	392	17178463264607761296016540993629780 705771520
124	388	177027055128131628050494707918077190 1717872640
128	384	1541987739885418045253212845850634832 46993999900
132	380	11380437366712812474455950864177326068 447989202944
136	376	713793445298874211607839796879716106185 715280216064
140	372	3816166003440131298948626476905412476595 9796671119360
144	368	17440779964066130420170168634612348393067 32612077058560
148	364	683209364930236126411369281492967750840643 659132148121 60
152	360	2299744204800465802453316637595783829108912 802028206751744
156	356	66674424868716978552789375387240003239187186 3497758510940 16
160	352	166855970096416058735080566458312292449892835 815171573300 7408
164	348	3611708227402789154515418737304813166113655239 003136470286 3360
168	344	6774835989895471077936151012477395142696211847 413560414611 04640
172	340	1103244193371309620166328638937318473011342162 120151575739708 2112
176	336	1562250954976198133076792319377808614268355671 567764765250841 77664
180	332	1926667532217097161576702991776654344250440175 688196887457279 508480
184	328	2072353402687653679228100239415179620504579373 643678880293833 6133120
188	324	1946714427418378529399755533637718562348412592 384043655562870 65292800
192	320	1599044990181340998819270766161596605692512085 057170791477694 075282632
196	316	1149841568524630218988847422278144249186012995 771486417325089 1967627264
200	312	7245946757074360381937881271877249754087077048 462649483895972 6267809792
204	308	4005499322639365542203429872582244997805641217 128274656743952 23861493760
208	304	1944071611978423909059426198144849863064608675 044397429548995 177751732480
212	300	8291211853278378544436157221213736835450108801 042695204524353 0869735424 00
216	296	3109550260070113076368271342789939024095055084 640910555058336 9693522427904
220	292	1026226524355102193549594379598979004344806158 459261421668544 26192158654464
224	288	2982062813021107266230007504454501325128818106 296071234784735 54095237810960
228	284	7633969196316666886767551069968038830038818474 387283118911093 846307975987 20
232	280	1722452776176219896357452486934573175804665343 735169479919087 8995825516871 68
236	276	3426750460257305904470547641506642175867699465 315478403354123 6313665086423 04
240	272	6013163599489683999312799935491777179772724247 998877953378442 9205014179338 24
244	268	9309551320248854051332692772889245412495562988 894547412532818 0450571164057 60
248	264	1271898604412951462071667415634190003046301502 177494040881598 9741288144568 320
252	260	1533699749994530538705635752791895045693439996 925023108607767 5815418680311 808
256		1632419990925168200043557728793436852309739769 254807177783748 3832108326674 502

References [1]

[1] C. Carlet and P. Solé, "The weight spectrum of two families of Reed-Muller codes", Discrete Mathematics, 346(10), 113568, 2023.

[2] P. Charpin, "Open problems on cyclic codes", in *Handbook of Coding Theory*, vol. 1, V. S. Pless, C. W. Huffman, editors, R. A. Brualdi, assistant editor, Elsevier, 963-1063, 1998.

[3] V. Gillot and Ph. Langevin, "Classification of some cosets of the Reed-Muller code", Cryptogr. Commun. (2023), available at https://doi.org/10.1007/s12095-023-00652-4.

[4] X. -D. Hou, "$GL(m, 2)$ acting on $\mathcal{R}(r, m)/\mathcal{R}(r-1, m)$", Discrete Mathematics, **149**, 99-122, 1996.

[5] A. Hulpke, "Computing with group orbits", available at https://www.math.colostate.edu/

[6] T. Kasami, N. Tokura, On the weight structure of Reed-Muller codes, IEEE Trans. Info. Theory, 16 , 752–759, 1970.

[7] T. Kasami, N. Tokura, S. Azumi, On the weight enumeration of weights less than 2.5d of Reed-Muller codes, Information and Control, 30, 380–395, 1976.

[8] Ph. Langevin, "Classification of Boolean quartic forms in eight variables", available at https://langevin.univ-tln.fr/project/quartics.html, 2007.

[9] Ph. Langevin, "Classification of $RM(4,7)/RM(2,7)$", available at https://langevin.univ-tln.fr/project/rm742/rm742.html, 2012.

[10] Ph. Langevin and G. Leander, "Classification of Boolean quartic forms in eight variables", in *Boolean Functions in Cryptology and Information Security*, B. Preneel and O. A. Logachev (Eds.), IOS Press, 139-147, 2008.

[11] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland Publishing Company, Amsterdam, New York, Oxford, 1977.

[12] D. V. Sarwate, *Weight Enumeratin of Reed-Muller Codes and Cosets*, Ph.D., Dep. Elec. Eng., Princeton Univ., Princeton, N.J., Sept. 1973, Advisors: E. R. Berlekamp and J. D. Ullman.

[13] N. J. A. Sloane, "On-line Encyclopedia of Integer Sequences" available at https://oeis.org/wiki/List of weight distributions.

THANKS FOR YOUR ATTENTION!