Correctable Errors of Weight Half the Minimum Distance Plus One for the First-Order Reed-Muller Codes

Kenji Yasunaga and Toru Fujiwara

Graduate School of Information Science and Technology,
Osaka University, Suita 565-0871, Japan
{k-yasunaga, fujiwara}@ist.osaka-u.ac.jp

Abstract. The number of correctable/uncorrectable errors of weight half the minimum distance plus one for the first-order Reed-Muller codes is determined. From a cryptographic viewpoint, this result immediately leads to the exact number of Boolean functions of m variables with nonlinearity $2^{m-2}+1$. The notion of *larger half* and *trial set*, which is introduced by Helleseth, Kløve, and Levenshtein to describe the monotone structure of correctable/uncorrectable errors, plays a significant role in the result.

Keywords: Syndrome decoding, Reed-Muller code, correctable error, Boolean function, nonlinearity, larger half.

1 Introduction

In syndrome decoding, the correctable errors are coset leaders of a code. The syndrome decoding performs maximum likelihood decoding if a minimum weight vector in each coset is taken as the coset leader. When there are two or more minimum weight vectors in a coset, we have choices of the coset leader. If the lexicographically smallest minimum weight vector is taken as the coset leader, then both the correctable errors and the uncorrectable errors have a monotone structure. That is, when $\boldsymbol{y}$ covers $\boldsymbol{x}$ (the support of $\boldsymbol{y}$ contains that of $\boldsymbol{x}$), if $\boldsymbol{y}$ is correctable, then $\boldsymbol{x}$ is also correctable, and if $\boldsymbol{x}$ is uncorrectable, then $\boldsymbol{y}$ is also uncorrectable [1]. Using this monotone structure, Helleseth, Kløve, and Levenshtein introduced *larger halves* of codewords and *trial sets* for codes to describe the monotone structure of errors and gave an improved upper bound on the number of uncorrectable errors using these notions [3].

The binary r-th order Reed-Muller code of length 2^m corresponds to the Boolean functions of m variables with degree at most r. The first-order Reed-Muller code of length 2^m, denoted by RM_m, corresponds to the set of affine functions of m variables. The *nonlinearity* of a Boolean function f of m variables is defined as the minimum distance between f and affine functions, and is equal to the weight of the coset leader in the coset f belongs to. Hence the weight distribution of coset leaders of RM_m represents the distribution of nonlinearity

S. Boztaş and H.F. Lu (Eds.): AAECC 2007, LNCS 4851, pp. 110–119, 2007.

of Boolean functions. When the number of coset leaders of weight i is p, the number of Boolean functions with the nonlinearity i is given by $p|\mathrm{RM}_m| = p2^{m+1}$. Nonlinearity is an important criterion for cryptographic system, in particular, block ciphers and stream ciphers. There has been much study of nonlinearity of Boolean functions in cryptography, see [4,5] and references therein. The weight distributions of the cosets of RM_5 are completely determined in [6]. In general, however, it is infeasible to obtain the weight distributions of the cosets (even only the coset leaders) of RM_m. Since the minimum distance of RM_m is 2^{m-1}, the problem is to know the number of the coset leaders of weight $\geq 2^{m-2}$. The explicit expression of the number of coset leaders of weight w, which is equal to the number of correctable errors of weight w, is given only for $w = 2^{m-2}$ [7].

In this paper, we determine the number of correctable/uncorrectable errors of weight $2^{m-2}+1$ for RM_m, from which the number of Boolean functions with nonlinearity $2^{m-2}+1$ is immediately obtained. To derive this result, we mainly use the properties of larger halves and trial sets.

2 Larger Halves and Trial Sets

Let $\mathbb{F}^n$ be the set of all binary vectors of length n. Let $C \subseteq \mathbb{F}^n$ be a binary linear code of length n, dimension k, and minimum distance d. Then $\mathbb{F}^n$ is partitioned into 2^{n-k} cosets $C_1, C_2, \ldots, C_{2^{n-k}}$; $\mathbb{F}^n = \bigcup_{i=1}^{2^{n-k}} C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$, where each $C_i = \{\boldsymbol{v}_i + \boldsymbol{c} : \boldsymbol{c} \in C\}$ with $\boldsymbol{v}_i \in \mathbb{F}^n$. The vector $\boldsymbol{v}_i$ is called the coset leader of the coset C_i, and any vector in C_i can be taken as $\boldsymbol{v}_i$.

Let H be a parity check matrix of C. The syndrome of a vector $\boldsymbol{v} \in \mathbb{F}^n$ is defined as $\boldsymbol{v}H^T$. All vectors having the same syndrome are in the same coset. Syndrome decoding associates an error vector to each syndrome. The syndrome decoder presumes that the error vector added to the received vector $\boldsymbol{y}$ is the coset leader of the coset which contains $\boldsymbol{y}$. The syndrome decoding function $D : \mathbb{F}^n \to C$ is defined as

$$D(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{v}_i \text{ if } \boldsymbol{y} \in C_i.$$

If each $\boldsymbol{v}_i$ has the minimum weight in its coset C_i, the syndrome decoder performs as a maximum likelihood decoder.

In this paper, we take as $\boldsymbol{v}_i$ the minimum element in C_i with respect to the following total ordering $\preceq$:

$$\boldsymbol{x} \preceq \boldsymbol{y} \text{ if and only if } \begin{cases} w(\boldsymbol{x}) < w(\boldsymbol{y}), & \text{or} \\ w(\boldsymbol{x}) = w(\boldsymbol{y}) \text{ and } v(\boldsymbol{x}) \leq v(\boldsymbol{y}), \end{cases}$$

where $w(\boldsymbol{x})$ denotes the Hamming weight of a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $v(\boldsymbol{x})$ denotes the numerical value of $\boldsymbol{x}$:

$$v(\boldsymbol{x}) = \sum_{i=1}^{n} x_i 2^{n-i}.$$

We write $\boldsymbol{x} \prec \boldsymbol{y}$ if $\boldsymbol{x} \preceq \boldsymbol{y}$ and $\boldsymbol{x} \neq \boldsymbol{y}$.

Let $E^0(C)$ be the set of all coset leaders of C. In the syndrome decoding, $E^0(C)$ is the set of correctable errors and $E^1(C) = \mathbb{F}^n \setminus E^0(C)$ is the set of uncorrectable errors. Since we take the minimum element with respect to $\preceq$ in each coset as its coset leader, both $E^0(C)$ and $E^1(C)$ have the following well-known monotone structure, see [1, Theorem 3.11]. Let $\subseteq$ denote a partial ordering called "covering" such that

$$\boldsymbol{x} \subseteq \boldsymbol{y} \text{ if and only if } S(\boldsymbol{x}) \subseteq S(\boldsymbol{y}),$$

where

$$S(\boldsymbol{v}) = \{i : v_i \neq 0\}$$

is the support of $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$. Consider $\boldsymbol{x}$ and $\boldsymbol{y}$ with $\boldsymbol{x} \subseteq \boldsymbol{y}$. If $\boldsymbol{y}$ is a correctable error, then $\boldsymbol{x}$ is also correctable. If $\boldsymbol{x}$ is uncorrectable, then $\boldsymbol{y}$ is also uncorrectable. For example, let $C = \{000, 001\}$ be a code. Then $E^0(C) = \{110, 100, 010\}$ and $E^1(C) = \{001, 101, 111\}$. In this case, even if we only know the fact that the vector 110 is correctable, we can deduce the vectors 100 and 010 are correctable, since they are covered by 110. A similar thing happens when we know 001 is uncorrectable. Using this structure, Zémor showed that the residual error probability after maximum likelihood decoding displays a threshold behavior [2]. Helleseth, Kløve, and Levenshtein [3] studied this structure and introduced *larger halves* and *trial sets*.

Since the set of uncorrectable errors $E^1(C)$ has a monotone structure, $E^1(C)$ can be characterized by *minimal uncorrectable errors* in $E^1(C)$. An uncorrectable error $\boldsymbol{y} \in E^1(C)$ is minimal if there exists no $\boldsymbol{x}$ such that $\boldsymbol{x} \subset \boldsymbol{y}$ in $E^1(C)$. If we know all minimal uncorrectable errors, all uncorrectable errors can be determined from them. We denote by $M^1(C)$ the set of all minimal uncorrectable errors in C. Larger halves of a codeword $\boldsymbol{c} \in C \setminus \{\boldsymbol{0}\}$ are introduced to characterize the minimal uncorrectable errors, and are defined as minimal vectors $\boldsymbol{v}$ with respect to covering such that $\boldsymbol{v} + \boldsymbol{c} \prec \boldsymbol{v}$. Any larger half $\boldsymbol{v}$ of a codeword $\boldsymbol{c}$ is an uncorrectable error, since $\boldsymbol{v} + \boldsymbol{c} \prec \boldsymbol{v}$ and they are in the same coset. The following condition is a necessary and sufficient condition that $\boldsymbol{v} \in \mathbb{F}^n$ is a larger half of $\boldsymbol{c} \in C \setminus \{\boldsymbol{0}\}$:

$$\boldsymbol{v} \subseteq \boldsymbol{c}, \tag{1}$$

$$w(\boldsymbol{c}) \leq 2w(\boldsymbol{v}) \leq w(\boldsymbol{c}) + 2, \tag{2}$$

$$l(\boldsymbol{v}) \begin{cases} = l(\boldsymbol{c}), & \text{if } \ 2w(\boldsymbol{v}) = w(\boldsymbol{c}), \\ > l(\boldsymbol{c}), & \text{if } \ 2w(\boldsymbol{v}) = w(\boldsymbol{c}) + 2, \end{cases} \tag{3}$$

where $l(\boldsymbol{x})$ is the smallest element in $S(\boldsymbol{x})$, that is, $l(\boldsymbol{x})$ is the leftmost non-zero coordinate in the vector $\boldsymbol{x}$. The proof of equivalence between the definition and the above condition is found in the proof of Theorem 1 of [3]. Let $LH(\boldsymbol{c})$ be the set of all larger halves of $\boldsymbol{c} \in C \setminus \{\boldsymbol{0}\}$. For a subset U of $C \setminus \{\boldsymbol{0}\}$, let

$$LH(U) = \bigcup_{\boldsymbol{c} \in U} LH(\boldsymbol{c}).$$

A trial set T for a code C is defined as follows:

$$T \subseteq C \setminus \{\mathbf{0}\} \text{ is a trial set for } C \text{ if } M^1(C) \subseteq LH(T). \tag{4}$$

A codeword $\boldsymbol{c}$ is called *minimal* if $\boldsymbol{c}' \subset \boldsymbol{c}$ for $\boldsymbol{c}' \in C$ implies $\boldsymbol{c}' = \mathbf{0}$. Let C^* be the set of all minimal codewords in C. It is shown that a trial set can consist of only minimal codewords [3, Corollary 5]. Therefore, C^* is a trial set of C.

In the rest of paper, for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^n$, we write $\boldsymbol{u} \cap \boldsymbol{v}$ as the vector in $\mathbb{F}^n$ whose support is $\mathrm{S}(\boldsymbol{u}) \cap \mathrm{S}(\boldsymbol{v})$.

3 Uncorrectable Errors of Weight $2^{m-2}+1$ for RM_m

In this section, we determine the number of correctable/uncorrectable errors of weight half the minimum distance plus one for the first-order Reed-Muller code of length $n = 2^m$, denoted by RM_m.

RM_m is a code of dimension $k = m + 1$, and minimum distance $d = 2^{m-1}$, and is defined recursively as

$$\mathrm{RM}_0 = \{\mathbf{0}, \mathbf{1}\},$$
$$\mathrm{RM}_m = \bigcup_{\boldsymbol{c} \in \mathrm{RM}_{m-1}} \{\boldsymbol{c} \circ \boldsymbol{c}, \boldsymbol{c} \circ \overline{\boldsymbol{c}}\},$$

where $\boldsymbol{u} \circ \boldsymbol{v}$ denotes the concatenation of $\boldsymbol{u}$ and $\boldsymbol{v}$, and $\overline{\boldsymbol{v}} \triangleq \mathbf{1} + \boldsymbol{v}$. Since all codewords in RM_m except all-zero and all-one codewords are minimum weight codewords, $\mathrm{RM}_m^* = \mathrm{RM}_m \setminus \{\mathbf{0}, \mathbf{1}\}$.

The weights of vectors in $LH(\mathrm{RM}_m^*)$ are 2^{m-2} and $2^{m-2}+1$ from the condition (2). Let $LH^-(\boldsymbol{c})$ and $LH^+(\boldsymbol{c})$ denote the sets of larger halves of $\boldsymbol{c} \in \mathrm{RM}_m^*$ of weight 2^{m-2} and $2^{m-2}+1$, respectively. Also let $LH^-(\mathrm{RM}_m^*) = \bigcup_{\boldsymbol{c} \in \mathrm{RM}_m^*} LH^-(\boldsymbol{c})$ and $LH^+(\mathrm{RM}_m^*) = \bigcup_{\boldsymbol{c} \in \mathrm{RM}_m^*} LH^+(\boldsymbol{c})$.

Let $E^1_{2^{m-2}+1}(\mathrm{RM}_m)$ be the set of uncorrectable errors of weight $d+1 = 2^{m-2}+1$ in RM_m. The set, $E^1_{2^{m-2}+1}(\mathrm{RM}_m)$, contains $LH^+(\mathrm{RM}_m^*)$, and $LH^+(\mathrm{RM}_m^*)$ contains all minimal uncorrectable errors of the weight from (4). Therefore, the remaining uncorrectable errors in $E^1_{2^{m-2}+1}(\mathrm{RM}_m)$ are non-minimal.

We will evaluate $|E^1_{2^{m-2}+1}(\mathrm{RM}_m)|$ by partitioning the set into two subsets. The first subset consists of the vectors that is covered by some codeword in RM_m^*. Any $\boldsymbol{v} \in \mathbb{F}^n$ of weight $2^{m-2}+1$ covered by $\boldsymbol{c} \in \mathrm{RM}_m$ is uncorrectable, since the coset to which $\boldsymbol{v}$ belongs contains the smaller weight vector $\boldsymbol{c} + \boldsymbol{v}$. The second one consists of the remaining non-minimal vectors.

Now, we evaluate the number of vectors in the first subset. It contains $\binom{2^{m-1}}{2^{m-2}+1}$ vectors for each codeword in RM_m^*, and all $|\mathrm{RM}_m^*| \cdot \binom{2^{m-1}}{2^{m-2}+1}$ such vectors are distinct. This is because, if $\boldsymbol{v} \subseteq \boldsymbol{c}_1$ and $\boldsymbol{v} \subseteq \boldsymbol{c}_2$ for a vector $\boldsymbol{v}$ in the set, then we have $w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2) \geq w(\boldsymbol{v}) = 2^{m-2}+1$, which contradicts the following Lemma 1.

Lemma 1. *Let $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathrm{RM}_m^*$ with $\boldsymbol{c}_1 \neq \boldsymbol{c}_2$. Then, it holds that*

$$w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2) = \begin{cases} 2^{m-2}, & \text{if } \boldsymbol{c}_1 + \boldsymbol{c}_2 \neq \mathbf{1}, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. The statement follows from the fact that $w(\boldsymbol{c}_1 + \boldsymbol{c}_2) = w(\boldsymbol{c}_1) + w(\boldsymbol{c}_2) - 2w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2)$. That is,

$$\begin{aligned} w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2) &= \frac{w(\boldsymbol{c}_1) + w(\boldsymbol{c}_2) - w(\boldsymbol{c}_1 + \boldsymbol{c}_2)}{2} \\ &= \frac{2^{m-1} + 2^{m-1} - w(\boldsymbol{c}_1 + \boldsymbol{c}_2)}{2} \\ &= \frac{2^m - w(\boldsymbol{c}_1 + \boldsymbol{c}_2)}{2}. \end{aligned}$$

□

Next, we evaluate the number of vectors in the second subset. The vectors in the subset are non-minimal uncorrectable errors that are not covered by any codeword in RM_m^*. Such a error covers a minimal uncorrectable error of weight 2^{m-2} in $LH^-(\mathrm{RM}_m^*)$, since 2^{m-2} is the smallest weight in uncorrectable errors. Therefore, we consider the set of vectors obtained by adding a weight-one vector to a larger half in $LH^-(\mathrm{RM}_m^*)$ that are not covered by any codeword in RM_m^*.

Let

$$\begin{aligned} &\mathbb{E}^n = \{\boldsymbol{e} \in \mathbb{F}^n : w(\boldsymbol{e}) = 1\}, \\ &\mathbb{E}^n(\boldsymbol{c}) = \{\boldsymbol{e} \in \mathbb{E}^n : \boldsymbol{e} \cap \boldsymbol{c} = \mathbf{0}\}, \quad \text{for } \boldsymbol{c} \in \mathrm{RM}_m^*. \end{aligned}$$

Then, the second subset can be represented as $\mathrm{X}_m \setminus \mathrm{Y}_m$, where

$$\begin{aligned} \mathrm{X}_m &= \{\boldsymbol{v} + \boldsymbol{e} : \boldsymbol{v} \in LH^-(\boldsymbol{c}) \text{ with } \boldsymbol{c} \in \mathrm{RM}_m^*, \boldsymbol{e} \in \mathbb{E}^n(\boldsymbol{c})\}, \\ \mathrm{Y}_m &= \{\boldsymbol{u} \in \mathrm{X}_m : \boldsymbol{u} \subseteq \boldsymbol{c} \text{ for some } \boldsymbol{c} \in \mathrm{RM}_m^*\}. \end{aligned}$$

From the above discussion, we have

$$|E^1_{2^{m-2}+1}(\mathrm{RM}_m)| = 2(2^m - 1)\binom{2^{m-1}}{2^{m-2}+1} + |\mathrm{X}_m \setminus \mathrm{Y}_m|. \tag{5}$$

For X_m and Y_m, we define the corresponding multisets $\tilde{\mathrm{X}}_m$ and $\tilde{\mathrm{Y}}_m$. That is, $\tilde{\mathrm{X}}_m$ is a multiset of vectors obtained by adding a weight-one vector $\boldsymbol{e}$ to larger halves $\boldsymbol{v} \in LH^-(\boldsymbol{c})$ satisfying $\boldsymbol{c} \cap \boldsymbol{e} = \mathbf{0}$ for each $\boldsymbol{c} \in \mathrm{RM}_m^*$. The set $\tilde{\mathrm{Y}}_m$ is a multiset of vectors in $\tilde{\mathrm{X}}_m$ that are covered by some codeword in RM_m^*. Then we have

$$\begin{aligned} |\tilde{\mathrm{X}}_m| &= |\mathrm{RM}_m^*| \cdot \binom{2^{m-1}-1}{2^{m-2}-1} \cdot 2^{m-1} \\ &= 2^{m-1}(2^m - 1)\binom{2^{m-1}}{2^{m-2}}, \end{aligned} \tag{6}$$

since the number of larger halves of each codeword is $\binom{2^{m-1}-1}{2^{m-2}-1}$ from (1)–(3).

We will evaluate $|\mathrm{X}_m \setminus \mathrm{Y}_m|$ by using $\tilde{\mathrm{X}}_m$ and $\tilde{\mathrm{Y}}_m$. First, we will show that the multiplicity of vectors in $\tilde{\mathrm{X}}_m \setminus \tilde{\mathrm{Y}}_m$ is not greater than 2 by using the following lemma.

Lemma 2. *Let $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ be distinct codewords in RM_m^*. Then it holds that*

$$w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3) = \begin{cases} 2^{m-2}, & \text{if } \boldsymbol{c}_1 + \boldsymbol{c}_2 + \boldsymbol{c}_3 = \mathbf{1}, \\ 0, & \text{if } \boldsymbol{c}_i + \boldsymbol{c}_j = \mathbf{1} \text{ for some } i, j \text{ with } 1 \le i \ne j \le 3, \\ 2^{m-3}, & \text{otherwise.} \end{cases}$$

Proof. The statement follows from the fact that $w(\boldsymbol{c}_1 + \boldsymbol{c}_2 + \boldsymbol{c}_3) = w(\boldsymbol{c}_1) + w(\boldsymbol{c}_2) + w(\boldsymbol{c}_3) - 2(w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2) + w(\boldsymbol{c}_2 \cap \boldsymbol{c}_3) + w(\boldsymbol{c}_1 \cap \boldsymbol{c}_3)) + 4w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3)$ and Lemma 1. □

From the lemma, we see that $w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3) = 2^{m-3}$ if and only if $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3, 1$ are linearly independent, that is, $a_1\boldsymbol{c}_1 + a_2\boldsymbol{c}_2 + a_3\boldsymbol{c}_3 + a_4\mathbf{1} = \mathbf{0}$ yields $a_1 = a_2 = a_3 = a_4 = 0$.

Lemma 3. *The multiplicity of any vector in $\tilde{\mathrm{X}}_m \setminus \tilde{\mathrm{Y}}_m$ is less than or equal to 2 for $m \ge 5$.*

Proof. Let $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$ be distinct codewords in RM_m^*. For $1 \le i \le 3$, suppose there exist $\boldsymbol{v}_i, \boldsymbol{e}_i, \boldsymbol{u}$ such that $\boldsymbol{v}_i \in LH^-(\boldsymbol{c}_i)$, $\boldsymbol{e}_i \in \mathbb{E}^n(\boldsymbol{c}_i)$, $\boldsymbol{u} = \boldsymbol{v}_i + \boldsymbol{e}_i$, and there exists no $\boldsymbol{c}_4 \in \mathrm{RM}_m^*$ satisfying $\boldsymbol{u} \subseteq \boldsymbol{c}_4$. First note that $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$, and $\mathbf{1}$ must be linearly independent for existing the above $\boldsymbol{v}_i, \boldsymbol{e}_i, \boldsymbol{u}$ for $1 \le i \le 3$ for $m \ge 4$.

If $\boldsymbol{v}_1 = \boldsymbol{v}_2$, then $\boldsymbol{v}_1 = \boldsymbol{c}_1 \cap \boldsymbol{c}_2 \subseteq \mathbf{1} + \boldsymbol{c}_1 + \boldsymbol{c}_2$ and $\boldsymbol{e}_1 = \boldsymbol{e}_2 \subseteq \mathbf{1} + \boldsymbol{c}_1 + \boldsymbol{c}_2$, and thus $\boldsymbol{v}_1 + \boldsymbol{e}_1 \subseteq \mathbf{1} + \boldsymbol{c}_1 + \boldsymbol{c}_2$, leading to the contradiction. Therefore $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ are distinct, and so are $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$. Then $w(\boldsymbol{v}_1 \cap \boldsymbol{v}_2 \cap \boldsymbol{v}_3) = 2^{m-2} - 2$, and thus $w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3) \ge w(\boldsymbol{v}_1 \cap \boldsymbol{v}_2 \cap \boldsymbol{v}_3) = 2^{m-2} - 2$. On the other hand, $w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3) = 2^{m-3}$ from Lemma 2. Thus we have $2^{m-3} \ge 2^{m-2} - 2$. The contradiction arises when $m \ge 5$. □

Thus, the size of $\mathrm{X}_m \setminus \mathrm{Y}_m$ is represented as follows.

$$|\mathrm{X}_m \setminus \mathrm{Y}_m| = |\tilde{\mathrm{X}}_m| - |\tilde{\mathrm{Y}}_m| - \frac{|\tilde{\mathrm{Z}}_m|}{2}, \tag{7}$$

where $\tilde{\mathrm{Z}}_m$ is the multiset defined as

$$\tilde{\mathrm{Z}}_m = \{\boldsymbol{v} \in \tilde{\mathrm{X}}_m : \boldsymbol{v} \not\subseteq \boldsymbol{c} \text{ for any } \boldsymbol{c} \in \mathrm{RM}_m^*, \text{ the multiplicity of } \boldsymbol{v} \text{ is } 2\}.$$

We will determine $|\tilde{\mathrm{Y}}_m|$ and $|\tilde{\mathrm{Z}}_m|$. The next lemma is useful to evaluate $|\tilde{\mathrm{Y}}_m|$.

Lemma 4. *Let $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathrm{RM}_m^*$. Then*

1. *there exist $\boldsymbol{v} \in LH^-(\boldsymbol{c}_1)$, $\boldsymbol{e} \in \mathbb{E}^n(\boldsymbol{c}_1)$ such that $\boldsymbol{v} + \boldsymbol{e} \subseteq \boldsymbol{c}_2$ if and only if*

$$\boldsymbol{c}_1 \ne \boldsymbol{c}_2 \quad \text{and} \quad l(\boldsymbol{c}_1) \in S(\boldsymbol{c}_2); \tag{8}$$

2. if (8) holds,

$$\{(\boldsymbol{v}, \boldsymbol{e}) : \boldsymbol{v} \in LH^{-}(\boldsymbol{c}_1), \boldsymbol{e} \in \mathbb{E}^n(\boldsymbol{c}_1), \boldsymbol{v} + \boldsymbol{e} \subseteq \boldsymbol{c}_2\} = \{(\boldsymbol{c}_1 \cap \boldsymbol{c}_2, \boldsymbol{e}) : \boldsymbol{e} \in \mathbb{E}^n, S(\boldsymbol{e}) \subseteq S(\boldsymbol{c}_2) \setminus S(\boldsymbol{c}_1)\}. \quad (9)$$

Proof. (First part) The only if part is obvious. We prove the if part. Let $\boldsymbol{v} = \boldsymbol{c}_1 \cap \boldsymbol{c}_2$. Since $\boldsymbol{c}_1 \neq \boldsymbol{c}_2$ and $\boldsymbol{c}_1 + \boldsymbol{c}_2 \neq \mathbf{1}$ from (8), we have $w(\boldsymbol{v}) = 2^{m-2}$ from Lemma 1. We have $l(\boldsymbol{v}) = l(\boldsymbol{c}_1)$ from $l(\boldsymbol{c}_1) \in S(\boldsymbol{c}_2)$. Thus $\boldsymbol{v} \in LH^{-}(\boldsymbol{c}_1)$. Clearly, we can take $\boldsymbol{e} \in \mathbb{E}^n(\boldsymbol{c}_1)$ such that $\boldsymbol{v} + \boldsymbol{e} \subseteq \boldsymbol{c}_2$.

(Second part) The $\supseteq$ part is obvious, so we show the $\subseteq$ part. Since $\boldsymbol{v} \subseteq \boldsymbol{c}_1$ and $\boldsymbol{v} \subseteq \boldsymbol{c}_2$, it holds $w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2) \geq w(\boldsymbol{v}) = 2^{m-2}$. On the other hand, $w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2) = 2^{m-2}$. Therefore we have $\boldsymbol{v} = \boldsymbol{c}_1 \cap \boldsymbol{c}_2$. It immediately follows that $S(\boldsymbol{e}) \subseteq S(\boldsymbol{c}_2) \setminus S(\boldsymbol{c}_1)$ from $\boldsymbol{c}_1 \cap \boldsymbol{e} = \mathbf{0}$ and $\boldsymbol{v} + \boldsymbol{e} \subseteq \boldsymbol{c}_2$. □

From Lemma 4, $\boldsymbol{v} + \boldsymbol{e} \in \tilde{\mathrm{X}}_m$ is covered by every $\boldsymbol{c}_2 \in \mathrm{RM}_m^*$ satisfying (8). The number of codewords $\boldsymbol{c}_2$ satisfying (8) is $|\mathrm{RM}_m|/2 - 2 = 2^m - 2$. There are $|S(\boldsymbol{c}_2) \setminus S(\boldsymbol{c}_1)| = 2^{m-2}$ choices of $\boldsymbol{e}$ from (9). Thus we have

$$\begin{aligned} |\tilde{\mathrm{Y}}_m| &= |\mathrm{RM}_m^*| \cdot (2^m - 2) \cdot 2^{m-2} \\ &= 2^m(2^m - 1)(2^{m-1} - 1). \end{aligned} \quad (10)$$

The following lemma is useful to derive $|\tilde{\mathrm{Z}}_m|$.

Lemma 5. *Let $\boldsymbol{u} \in \tilde{\mathrm{X}}_m$ of multiplicity 2. That is, $\boldsymbol{u}$ is represented as $\boldsymbol{u} = \boldsymbol{v}_1 + \boldsymbol{e}_1 = \boldsymbol{v}_2 + \boldsymbol{e}_2$ where $\boldsymbol{v}_i \in LH^{-}(\boldsymbol{c}_i)$, $\boldsymbol{c}_i \in \mathrm{RM}_m^*$, $\boldsymbol{e}_i \in \mathbb{E}^n(\boldsymbol{c}_i)$ for $i = 1, 2$, and $\boldsymbol{c}_1 \neq \boldsymbol{c}_2$. Then,*

1. *for $m \geq 3$, $\boldsymbol{c}_1 + \boldsymbol{c}_2 \neq \mathbf{1}$,*
2. *for $m \geq 5$, there exists $\boldsymbol{c}_3 \in \mathrm{RM}_m^*$ such that $\boldsymbol{u} \subseteq \boldsymbol{c}_3$ if and only if $\boldsymbol{e}_1 = \boldsymbol{e}_2$.*

Proof. The first part holds, since $\boldsymbol{v}_1 + \boldsymbol{e}_1 = \boldsymbol{v}_2 + \boldsymbol{e}_2$ cannot hold for $m \geq 3$ if $\boldsymbol{c}_1 + \boldsymbol{c}_2 = \mathbf{1}$. Now we prove the second part.

(Only if part) We have $\boldsymbol{c}_1 \neq \boldsymbol{c}_3$ from $\boldsymbol{v}_1 + \boldsymbol{e}_1 \not\subseteq \boldsymbol{c}_1$ and $\boldsymbol{v}_1 + \boldsymbol{e}_1 \subseteq \boldsymbol{c}_3$. Since $\boldsymbol{v}_1 \subseteq \boldsymbol{c}_1$, and $\boldsymbol{v}_1 \subseteq \boldsymbol{c}_3$, we have $\boldsymbol{v}_1 = \boldsymbol{c}_1 \cap \boldsymbol{c}_3$. Equivalently, $\boldsymbol{v}_2 = \boldsymbol{c}_2 \cap \boldsymbol{c}_3$. Then $\boldsymbol{v}_1 \cap \boldsymbol{v}_2 = \boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3$, and hence $w(\boldsymbol{v}_1 \cap \boldsymbol{v}_2) = w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3)$. Since $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$ are distinct, $w(\boldsymbol{c}_1 \cap \boldsymbol{c}_2 \cap \boldsymbol{c}_3)$ is either 2^{m-2}, 2^{m-3}, or 0. On the other hand, $w(\boldsymbol{v}_1 \cap \boldsymbol{v}_2)$ is 2^{m-2} if $\boldsymbol{v}_1 = \boldsymbol{v}_2$, and is $2^{m-2} - 2$ otherwise, since $\boldsymbol{v}_1 + \boldsymbol{e}_1 = \boldsymbol{v}_2 + \boldsymbol{e}_2$. Therefore $w(\boldsymbol{v}_1 \cap \boldsymbol{v}_2) = 2^{m-2}$ for $m \geq 5$, since $2^{m-3} \neq 2^{m-2} - 2$. Hence $\boldsymbol{v}_1 = \boldsymbol{v}_2$, and thus $\boldsymbol{e}_1 = \boldsymbol{e}_2$.

(If part) Since $\boldsymbol{e}_1 = \boldsymbol{e}_2$ and $\boldsymbol{c}_1 \neq \boldsymbol{c}_2$, we have $\boldsymbol{v}_1 = \boldsymbol{v}_2 = \boldsymbol{c}_1 \cap \boldsymbol{c}_2 \subseteq \mathbf{1} + \boldsymbol{c}_1 + \boldsymbol{c}_2$. Since $\boldsymbol{e}_1 \cap \boldsymbol{c}_1 = \boldsymbol{e}_2 \cap \boldsymbol{c}_2 = \boldsymbol{e}_1 \cap \boldsymbol{c}_2 = \mathbf{0}$, we have $\boldsymbol{e}_1 \subseteq \mathbf{1} + \boldsymbol{c}_1 + \boldsymbol{c}_2$. By taking $\boldsymbol{c}_3 = \mathbf{1} + \boldsymbol{c}_1 + \boldsymbol{c}_2$, we have $\boldsymbol{u} = \boldsymbol{v}_1 + \boldsymbol{e}_1 \subseteq \boldsymbol{c}_3$. □

From Lemma 5, for each $\boldsymbol{c}_1 \in \mathrm{RM}_m^*$, $|\tilde{\mathrm{Z}}_m|$ is obtained by counting all patterns in $\{\boldsymbol{v}_1 + \boldsymbol{e}_1 : \boldsymbol{v}_1 \in LH^{-}(\boldsymbol{c}_1), \boldsymbol{e}_1 \in \mathbb{E}^n(\boldsymbol{c}_1)\}$ such that $\boldsymbol{v}_1 + \boldsymbol{e}_1 = \boldsymbol{v}_2 + \boldsymbol{e}_2$ for some

$\boldsymbol{v}_2, \boldsymbol{e}_2$ with $\boldsymbol{v}_2 \in LH^-(\boldsymbol{c}_2), \boldsymbol{c}_2 \in \mathrm{RM}_m^* \setminus \{\boldsymbol{c}_1\}, \boldsymbol{e}_2 \in \mathbb{E}^n(\boldsymbol{c}_2)$ and $\boldsymbol{e}_1 \neq \boldsymbol{e}_2$. We will count such $\boldsymbol{v}_1 + \boldsymbol{e}_1$ for each $\boldsymbol{c}_1 \in \mathrm{RM}_m^*$.

We introduce some notations. Let $S_m = \{l(\boldsymbol{c}) : \boldsymbol{c} \in \mathrm{RM}_m\}$. From the definition of RM_m, $S_m = \{s_1, s_2, \ldots, s_k\}$, where

$$s_i = \begin{cases} 1, & \text{for } i = 1, \\ 2^{i-2} + 1, & \text{for } 2 \leq i \leq k = m + 1. \end{cases}$$

Also define

$$C_m(s_i) = \{\boldsymbol{c} \in \mathrm{RM}_m^* : l(\boldsymbol{c}) = s_i\}.$$

Then, we have

$$|C_m(s_i)| = \begin{cases} 2^m - 1, & \text{for } i = 1, \\ 2^{m+1-i}, & \text{for } 2 \leq i \leq m + 1. \end{cases} \tag{11}$$

Now we are ready to evaluate $|\tilde{\mathrm{Z}}_m|$. There are three cases to be considered.

1. When $l(\boldsymbol{c}_1) = l(\boldsymbol{c}_2)$; we choose $\boldsymbol{w}$ such that $\boldsymbol{w} \subseteq \boldsymbol{c}_1 \cap \boldsymbol{c}_2, w(\boldsymbol{w}) = 2^{m-2} - 1$, and $l(\boldsymbol{w}) = l(\boldsymbol{c}_1 \cap \boldsymbol{c}_2)$. We choose $\boldsymbol{e}_2$ so that $S(\boldsymbol{e}_2) \subseteq S(\boldsymbol{c}_1) \setminus S(\boldsymbol{c}_2)$, and choose $\boldsymbol{e}_1$ so that $S(\boldsymbol{e}_1) \subseteq S(\boldsymbol{c}_2) \setminus S(\boldsymbol{c}_1)$. Then letting $\boldsymbol{v}_1 = \boldsymbol{w} + \boldsymbol{e}_2$ and $\boldsymbol{v}_2 = \boldsymbol{w} + \boldsymbol{e}_1$ gives vectors as $\boldsymbol{v}_1 + \boldsymbol{e}_1 = \boldsymbol{v}_2 + \boldsymbol{e}_2$. There are $(2^{m-2} - 1) \cdot 2^{m-2} \cdot 2^{m-2}$ such $\boldsymbol{v}_1 + \boldsymbol{e}_1$.

   For each codeword $\boldsymbol{c}_1$ in $C_m(s_i)$, there are $|C_m(s_i)| - 1$ codewords $\boldsymbol{c}_2$ in RM_m^* satisfying $l(\boldsymbol{c}_1) = l(\boldsymbol{c}_2)$.
2. When $l(\boldsymbol{c}_1) > l(\boldsymbol{c}_2)$; since $\boldsymbol{v}_1 \in LH^-(\boldsymbol{c}_1)$ and $\boldsymbol{v}_2 \in LH^-(\boldsymbol{c}_2)$, the $l(\boldsymbol{c}_2)$-th bit of $\boldsymbol{e}_1$ is one.

   (a) If the $l(\boldsymbol{c}_1)$-th bit of $\boldsymbol{c}_2$ is one; we choose $\boldsymbol{w}$ such that $\boldsymbol{w} \subseteq \boldsymbol{c}_1 \cap \boldsymbol{c}_2, w(\boldsymbol{w}) = 2^{m-2} - 1$, and $l(\boldsymbol{w}) = l(\boldsymbol{c}_1 \cap \boldsymbol{c}_2)$. We choose $\boldsymbol{e}_2$ so that $S(\boldsymbol{e}_2) \subseteq S(\boldsymbol{c}_1) \setminus S(\boldsymbol{c}_2)$. Then letting $\boldsymbol{v}_1 = \boldsymbol{w} + \boldsymbol{e}_2$ and $\boldsymbol{v}_2 = \boldsymbol{w} + \boldsymbol{e}_1$ gives vectors as $\boldsymbol{v}_1 + \boldsymbol{e}_1 = \boldsymbol{v}_2 + \boldsymbol{e}_2$. There are $(2^{m-2} - 1) \cdot 2^{m-2}$ such $\boldsymbol{v}_1 + \boldsymbol{e}_1$.

   For each codeword $\boldsymbol{c}_1$ in $C_m(s_i)$ with $i \geq 2$, there are $\left(\left(\sum_{j<i} |C_m(s_j)| + 1\right)/2 - 1\right)$ codewords $\boldsymbol{c}_2$ in RM_m^* satisfying $l(\boldsymbol{c}_1) \in S(\boldsymbol{c}_2)$.

   (b) If the $l(\boldsymbol{c}_1)$-th bit of $\boldsymbol{c}_2$ is zero; then $\boldsymbol{e}_2$ must be the vector having one in the $l(\boldsymbol{c}_1)$-th bit. We choose $\boldsymbol{w}$ such that $\boldsymbol{w} \subseteq \boldsymbol{c}_1 \cap \boldsymbol{c}_2$ and $w(\boldsymbol{w}) = 2^{m-2} - 1$. Then letting $\boldsymbol{v}_1 = \boldsymbol{w} + \boldsymbol{e}_2$ and $\boldsymbol{v}_2 = \boldsymbol{w} + \boldsymbol{e}_1$ gives vectors as $\boldsymbol{v}_1 + \boldsymbol{e}_1 = \boldsymbol{v}_2 + \boldsymbol{e}_2$. There are 2^{m-2} such $\boldsymbol{v}_1 + \boldsymbol{e}_1$.

   For each codeword $\boldsymbol{c}_1$ in $C_m(s_i)$ with $i \geq 2$, there are $\left(\left(\sum_{j<i} |C_m(s_j)| + 1\right)/2 - 1\right)$ codewords $\boldsymbol{c}_2$ in RM_m^* satisfying $l(\boldsymbol{c}_1) \notin S(\boldsymbol{c}_2)$ and $\boldsymbol{c}_1 + \boldsymbol{c}_2 \neq \mathbf{1}$.
3. When $l(\boldsymbol{c}_1) < l(\boldsymbol{c}_2)$; the number of vectors we should count is equal to that for the second case.

From the above analysis, we have

$$\begin{aligned}|\tilde{\mathrm{Z}}_m| &= \sum_{i=1}^{m+1} |C_m(s_i)|(|C_m(s_i)|-1)\cdot(2^{m-2}-1)(2^{m-2})^2 \\ &\quad + 2\sum_{i=2}^{m+1} |C_m(s_i)| \left(\left(\sum_{j=1}^{i-1}|C_m(s_j)|+1\right)\cdot\frac{1}{2}-1\right)\cdot(2^{m-2}-1)2^{m-2} \\ &\quad + 2\sum_{i=2}^{m+1} |C_m(s_i)| \left(\left(\sum_{j=1}^{i-1}|C_m(s_j)|+1\right)\cdot\frac{1}{2}-1\right)\cdot 2^{m-2} \\ &= 2^{2m-3}\binom{2^m}{3}. \end{aligned} \tag{12}$$

From (5), (6), (7), (10), and (12), we can determine the number of uncorrectable errors of weight $2^{m-2}+1$ for RM_m.

Theorem 1. *For $m \geq 5$,*

$$|E^1_{2^{m-2}+1}(\mathrm{RM}_m)| = 4(2^m-1)(2^{m-3}+1)\binom{2^{m-1}}{2^{m-2}+1} - (4^{m-2}+3)\binom{2^m}{3}.$$

The number of correctable errors of weight $2^{m-2}+1$, $|E^0_{2^{m-2}+1}(\mathrm{RM}_m)|$, is obtained from the equation,

$$|E^0_{2^{m-2}+1}(\mathrm{RM}_m)| + |E^1_{2^{m-2}+1}(\mathrm{RM}_m)| = \binom{2^m}{2^{m-2}+1}.$$

On the number of Boolean functions with m variables of nonlinearity $2^{m-2}+1$, we have Corollary 1.

Corollary 1. *For $m \geq 5$, the number of Boolean functions with m variables of nonlinearity $2^{m-2}+1$ is equal to $2^{m+1}|E^0_{2^{m-2}+1}(\mathrm{RM}_m)|$, which is*

$$2^{m+1}\left(\binom{2^m}{2^{m-2}+1} - 4(2^m-1)(2^{m-3}+1)\binom{2^{m-1}}{2^{m-2}+1} + (4^{m-2}+3)\binom{2^m}{3}\right).$$

The results of the calculation of $|E^0_{2^{m-2}+1}(\mathrm{RM}_m)|$ and $|E^1_{2^{m-2}+1}(\mathrm{RM}_m)|$ for $5 \leq m \leq 9$ are listed in Table 1. These expressions can be approximated by Stirling's approximation, $n! \approx \sqrt{2\pi n}(n/e)^n$. Thereby,

$$|E^0_{2^{m-2}+1}(\mathrm{RM}_m)| \approx \sqrt{\frac{3}{2^{m-3}\pi}}\left(\frac{16}{3\sqrt{3}}\right)^{2^{m-1}},$$

$$|E^1_{2^{m-2}+1}(\mathrm{RM}_m)| \approx \sqrt{\frac{2^m}{\pi}}(2^m+8)2^{2^{m-1}}.$$

The ratio, $|E^1_{2^{m-2}+1}(\mathrm{RM}_m)|/|E^0_{2^{m-2}+1}(\mathrm{RM}_m)|$, approaches zero as m increases.

Table 1. The number of correctable/uncorrectable errors of weight $2^{m-2}+1$ for RM_m

			correctable	uncorrectable
m	n	k	$\lvert E^0_{2^{m-2}+1}(\mathrm{RM}_m)\rvert$	$\lvert E^1_{2^{m-2}+1}(\mathrm{RM}_m)\rvert$
5	32	6	21,288,320	6,760,480
6	64	7	1.378×10^{15}	1.283×10^{12}
7	128	8	4.299×10^{30}	1.535×10^{22}
8	256	9	5.625×10^{61}	7.938×10^{41}
9	512	10	1.329×10^{124}	7.605×10^{80}

4 Conclusion

In this paper, we have determined the number of correctable/uncorrectable errors of weight half the minimum distance plus one for the first-order Reed-Muller codes. We mainly use the notion of larger halves to derive this result.

Future work includes deriving the number of correctable errors of weight $\geq 2^{m-2}+2$ for RM_m using the larger half technique and applying the technique to other codes, for example, the second-order Reed-Muller codes and BCH codes.

References

1. Peterson, W.W., Weldon Jr., E.J.: Error-Correcting Codes, 2nd edn. MIT Press, Cambridge (1972)
2. Zémor, G.: Threshold Effects in Codes. In: Cohen, G., Lobstein, A., Zémor, G., Litsyn, S.N. (eds.) Algebraic Coding. LNCS, vol. 781, pp. 278–286. Springer, Heidelberg (1994)
3. Helleseth, T., Kløve, T., Levenshtein, V.: Error-Correction Capability of Binary Linear Codes. IEEE Trans. Infom. Theory 51(4), 1408–1423 (2005)
4. Canteaut, A., Carlet, C., Charpin, P., Fontaine, C.: On Cryptographic Properties of the Cosets of $R(1, m)$. IEEE Trans. Inform. Theory 47(4), 1513–1949 (2001)
5. Carlet, C.: Boolean Functions for Cryptography and Error Correcting Codes. In: Crama, Y., Hammer, P. (eds.) Boolean Methods and Models, Cambridge University Press, Cambridge (press)
6. Berlekamp, E.R., Welch, L.R.: Weight Distributions of the Cosets of the (32,6) Reed-Muller Code. IEEE Trans. Inform. Theory 18(1), 203–207 (1972)
7. Wu, C.K.: On Distribution of Boolean Functions with Nonlinearity $\leq 2^{n-2}$: Australasian. Journal of Combinatorics 17, 51–59 (1998)