Counting Balanced Boolean Functions in n Variables with Bounded Degree

Thomas W. Cusick and Younhwan Cheon

CONTENTS

We consider the problem of obtaining good upper and lower bounds on the number of balanced Boolean functions in n variables with degree less than or equal to k. This is the same as the problem of finding bounds on the number of codewords of weight 2^{n-1} in the Reed–Muller code of length 2^n and order k. We state several conjectures and use them to obtain good bounds. We believe that the conjectures will be highly useful for further research.

1. INTRODUCTION

The goal of this paper is to use some plausible conjectures to derive good upper and lower bounds on the number $B(k,n)$, which is defined as the number of balanced Boolean functions in n variables with degree less than or equal to k. As usual, we say that a Boolean function in n variables is balanced if exactly half of its 2^n values are zero.

It is obvious that

$$B(n,n) = \binom{2^n}{2^{n-1}},$$

and we shall see later that $B(n-1,n) = B(n,n)$. There is a formula for $B(n-2,n)$ that goes back at least as far as [Camion 79, p. 8], but it is doubtful that there is any simple formula for $B(k,n)$ for $k < n-2$. One reason for this doubt is the connection of $B(k,n)$ with the notorious weight-distribution problem for Reed–Muller codes.

An excellent account of the theory of Reed–Muller codes is given in [MacWilliams and Sloane 78, Chapters 13–15]. The Reed–Muller codes are defined in terms of Boolean functions as follows: Any Boolean function $f(x) = f(x_1, \ldots, x_n)$ can be uniquely specified by a truth table that lists the 2^n different values of $f(x)$ as x varies. We assume that x runs through its values in lexicographical order, beginning with $(0,\ldots,0)$ and $(0,\ldots,0,1)$ and

2000 AMS Subject Classification: Primary 94C10; Secondary 94B65, 11T71

Keywords: Boolean functions, balanced, Reed-Muller codes

ending with $(1, \ldots, 1)$. Thus for example, the truth table for

$$f(x_1, x_2, x_3) = x_1 + x_3 + x_1x_2 + x_1x_2x_3 \tag{1–1}$$

would be given by

$$(0, 1, 0, 1, 1, 0, 0, 0).$$

We will identify a Boolean function with its truth table, so we could write

$$f(x) = (0, 1, 0, 1, 1, 0, 0, 0)$$

instead of (1–1). The kth-order Reed–Muller code of length 2^n, denoted by $R(k, n)$, is the set of all vectors (truth tables) associated with Boolean functions $f(x)$ in n variables with degree less than or equal to k. These vectors are called the codewords of $R(k, n)$.

The weight of a Boolean function $f(x)$, denoted by $\mathrm{wt}(f)$, is the number of 1's in the truth table for the function f, so $\mathrm{wt}(f)$ is simply the Hamming weight of the vector associated with f. The weight distribution of $R(k, n)$ is the list of the possible weights of the codewords in $R(k, n)$, along with a count of how many times each weight occurs.

The weight distribution of $R(1, n)$ is trivial, and the weight distribution of $R(2, n)$ is known (see [MacWilliams and Sloane 78, pp. 434–445]). From these one can deduce by duality the weight distribution of $R(n - 2, n)$ and $R(n - 3, n)$, but in other cases there is no known formula for the weight distribution of any general class of Reed–Muller codes. Even the case $R(3, n)$ remains mysterious.

As far back as 1970 some progress was made in counting the codewords of small or large weight in $R(k, n)$ [MacWilliams and Sloane 78, Chapter 15, Theorem 11, p. 446]. There has been no significant progress for codewords of middling weight since then. Since $B(k, n)$ counts the codewords of the middle weight 2^{n-1}, we cannot expect a simple formula. In fact, it seems difficult even to prove a good estimate for $B(k, n)$.

As explained below, if we make some plausible conjectures, then we can achieve good upper and lower bounds on $B(k, n)$. Even if these conjectures cannot be proved, we can use them to derive interesting consequences, which can then possibly be proved or at least tested for correctness. There is precedent for the usefulness of difficult but insightful conjectures (for example, the Riemann hypothesis and Schanuel's conjecture).

2. A BALANCED FUNCTION DISTRIBUTION CONJECTURE

Let $F(k, n)$ denote the set of all Boolean functions in n variables of degree less than or equal to k. Let $P(k, n) = |F(k, n)|$, where as usual, $|S|$ denotes the number of elements in the set S. Hence

$$P(k, n) = 2^{1+C(n,1)+C(n,2)+\cdots+C(n,k)}, \tag{2–1}$$

where $C(n, j) = \binom{n}{j}$, $0 \le j \le n$.

Let $G(k, n)$ denote the set of all Boolean functions in n variables that are made up of one or more terms of exact degree k, so

$$|G(k, n)| = 2^{C(n,k)}. \tag{2–2}$$

We make the following conjecture:

Conjecture 2.1. *For any function g in $G(k+1, n)$, $k > 0$, we have*

$$\left|\{\, g + f : f \text{ runs through } F(k, n) \text{ and } g + f \text{ is balanced} \,\}\right| < B(k, n).$$

That is, the coset $R(k, n)$ in $R(k+1, n)/R(k, n)$ has more balanced functions than any other coset.

We believe that this conjecture gives a fundamental property of the cosets $R(k, n)$. We give two plausible consequences of Conjecture 2.1. To distinguish these from results that we can prove, we explicitly state that Conjecture 2.1 must be assumed.

Corollary 2.2. (Assume Conjecture 2.1.) *For each k, $1 \le k \le n - 1$, we have*

$$B(k, n) \cdot 2^{C(n,k+1)} \; > \; B(k + 1, n).$$

Proof: This follows immediately from (2–1) and Conjecture 2.1. □

Corollary 2.3. (Assume Conjecture 2.1.) *For each k, $1 \le k \le n - 1$, we have*

$$\frac{B(k + 1, n)}{P(k + 1, n)} < \frac{B(k, n)}{P(k, n)}.$$

Proof: Since

$$2^{C(n,k+1)} B(k, n)/P(k + 1, n) = B(k, n)/P(k, n),$$

the result follows from Corollary 2.2. □

Coset Representative f_i	Balanced Functions in $f_i + R(2,6)$
0	1,828,134
$x_1x_2x_3$	1,702,624
$x_1x_2x_3 + x_2x_4x_5$	1,675,520
$x_1x_2x_3 + x_4x_5x_6$	0
$x_1x_2x_3 + x_2x_4x_5 + x_3x_4x_6$	1,665,664
$x_1x_2x_3 + x_1x_4x_5 + x_2x_4x_6 + x_3x_5x_6 + x_4x_5x_6$	0

TABLE 1. Balanced functions in the cosets of $R(3,6)/R(2,6)$.

Coset Representative f_i	Balanced Functions in $f_i + R(2,7)$
0	300,503,590
$x_1x_2x_3$	171,425,760
$(x_1x_2 + x_3x_4)x_5$	156,664,832
$x_1x_2x_3 + x_4x_5x_6$	152,199,168
$(x_1x_2 + x_3x_4)x_5 + x_1x_3x_6$	153,664,128
$x_1x_2x_3 + x_4x_5x_6 + (x_1 + x_4)(x_2 + x_5)(x_3 + x_6)$	151,797,760
$(x_1x_2 + x_3x_4 + x_5x_6)x_7$	152,985,600
$(x_1x_2 + x_3x_4)x_5 + x_1x_6x_7$	151,417,600
$x_1x_2x_3 + x_4x_5x_6 + (x_1 + x_4)(x_2 + x_5)x_7$	151,169,024
$(x_1x_2 + x_3x_4)x_5 + x_1x_3x_6 + x_2x_3x_7$	152,182,784
$x_1x_2x_3 + x_4x_5x_6 + (x_1 + x_4)(x_2 + x_5)(x_3 + x_6) + x_1x_4x_7$	151,080,960
$x_1x_2x_3 + x_4x_5x_6 + x_2x_3 + x_4x_6 + ((x_1 + x_4)(x_2 + x_5) + x_3x_6)x_7$	151,057,408

TABLE 2. Balanced functions in the cosets of $R(3,7)/R(2,7)$.

The large computations necessary to determine the weight distribution of all cosets in $R(k+1,n)/R(k,n)$ have been carried out for $k = 2, n = 6$ (in [Hou 96]) and for $k = 2, n = 7$ (in [Kasami et al. 95]). In the $n = 6$ case, there are only 6 cosets $f_i + R(2,6)$ that have different weight distributions. In the $n = 7$ case, there are only 12 cosets $f_i + R(2,7)$ that have different weight distributions. Tables 1 and 2 give coset representatives f_i in these two cases, plus a count of the number of balanced functions in each coset. Conjecture 2.1 is confirmed in both cases. It is interesting to note the very uneven distribution of the balanced functions in the cosets of $R(2,6)$ (where two cosets have no balanced functions at all), as contrasted with the rather even distribution of the balanced functions in the cosets of $R(2,7)$ other than coset $R(2,7)$ itself.

The following deep theorem of McEliece (see [McEliece 72] or [MacWilliams and Sloane 78, p. 447]) motivates our Conjectures 3.2 and 4.1, which can be thought of as attempts to make more precise the role played by the powers of 2. Here $[\cdot]$ is the usual greatest-integer function.

Theorem 2.4. *The weight of every codeword in $R(k,n)$ is divisible by $2^{[(n-1)/k]}$.*

We cannot prove Conjecture 2.1 in general, but we can prove the special case $k = n-1$ (which is well known in coding theory):

Theorem 2.5. *For each $n \geq 2$, the set $x_1x_2 \cdots x_n + F(n-1,n)$ has no balanced functions; that is, $B(n-1,n) = B(n,n)$.*

Proof: Let $f_n(x)$ denote any function in $F(n-1,n)$. It suffices to show that

$$\sum_{x \in GF(2)^k} f_k(x) \equiv 0 \pmod 2 \qquad (2\text{–}3)$$

for each $k = 2, 3, \ldots,$ since then the truth table of $x_1x_2 \cdots x_k + f_k(x)$ will have an odd number of 1's. This simply means that all of the codewords in $R(n-1,n)$ have even weight, and this follows from the trivial weight distribution of $R(n-1,n)$. □

3. LOWER BOUNDS FOR $B(k,n)$

We quickly get a lower bound for $B(k,n)$ from Corollary 2.2 of Conjecture 2.1.

Theorem 3.1. (Assume Conjecture 2.1.) *For $1 \leq k \leq n-1$, we have*

$$B(k,n) > \binom{2^n}{2^{n-1}} \frac{P(k,n)}{2^{2^n-1}}.$$

Proof: By Corollary 2.3 we have

$$B(k,n) > \frac{P(k,n)B(n-1,n)}{P(n-1,n)},$$

Lower Bound	$\boldsymbol{B(k,n)}$	**Upper Bound**
12,870	$B(3,4) = 12{,}870$	12,870
18,783,800	$B(3,5) = 18{,}796{,}230^a$	1.89311×10^7
8.73863×10^{11}	$B(3,6) = 874{,}731{,}154{,}374^b$	8.77283×10^{11}
$2.863475220222 \times 10^{16}$	$B(4,6) = 2.86347527939 \times 10^{16\,a}$	2.874682×10^{16}
$5.193576914 \times 10^{18}$	$B(3,7) = 5.193595576 \times 10^{18\,c}$	5.20373×10^{18}
$8.922497198703 \times 10^{28}$	$B(4,7) = 8.922497198992 \times 10^{28\,a}$	8.939940×10^{28}
$1.973538269 \times 10^{27}$	$B(3,8) = 1.973540804 \times 10^{27\;\;d}$	1.97547×10^{27}
$1.164971371872 \times 10^{48}$	$B(4,8) = 1.164971371906 \times 10^{48\;\;a}$	$1.16610959 \times 10^{48}$
$1.91890047 \times 10^{38}$	$B(3,9) = 1.9189023 \times 10^{38\;\;e}$	1.91984×10^{38}
6.94355×10^{113}	$B(5,9) = 6.94355 \times 10^{113}$	6.94694×10^{113}
$1.910087568 \times 10^{52}$	$B(3,10) = 1.9100875806 \times 10^{52f}$	1.91055×10^{52}

Key: *a*, [Koumoto 04]; *b*, [Hou 96, Koumoto 04]; *c*, [Koumoto 04, Sugino et al. 71]; *d*, [Hou 94, **?**]; *e*, [Koumoto 04, Sugita et al. 96]; *f*, [Langevin 03].

TABLE 3. Number of balanced Boolean functions.

and by Theorem 2.5,

$$B(n-1,n) = B(n,n) = \binom{2^n}{2^{n-1}}.$$

Theorem 3.1 now follows from (2–1). □

We obtain a stronger lower bound if we assume the following stronger version of Corollary 2.3.

Conjecture 3.2. *If $3 \le k \le n-1$, then*

$$2^{[(n-1)/k]}\frac{B(k+1,n)}{P(k+1,n)} \le 2^{[(n-1)/(k+1)]}\frac{B(k,n)}{P(k,n)}. \qquad (3\text{–}1)$$

We note that equality holds in Conjecture 3.2 when $k = n-1$. Also, Conjecture 3.2 is the same as Corollary 2.3 with equality allowed except where $n \equiv 1 \pmod{k}$. Using Conjecture 3.2 we get the following improvement of the lower bound for $B(k,n)$.

Theorem 3.3. (Assume Conjecture 3.2). *For $3 \le k \le n-1$, we have*

$$B(k,n) \ge 2^{[(n-1)/k]}\binom{2^n}{2^{n-1}}\frac{P(k,n)}{2^{2^n}}. \qquad (3\text{–}2)$$

Proof: Applying (3–1) successively for $k, k+1, \ldots, n-1$ we obtain

$$\frac{B(k,n)}{P(k,n)} \ge 2^{[(n-1)/k]-1}\frac{B(n-1,n)}{P(n-1,n)}.$$

Now the theorem follows from

$$B(n-1,n) = \frac{P(n-1,n)}{2^{2^n-1}}\binom{2^n}{2^{n-1}},$$

which is just another way of stating Theorem 2.5. □

Note that for $k = n-1$ equality holds in (3–2). Table 3 shows that the lower bound is very good for the known values of $B(k,n)$. For example, the lower bound for $B(5,9)$ agrees with the actual value for the first 34 most-significant digits.

From [Sugita et al. 96] we have the lower bound

$$6.9435467290134140225804930314175505305 6 \times 10^{113}$$

for $B(5,9)$, and the exact value

$$6.94354672901341402258049303141755061046 \times 10^{113}.$$

4. UPPER BOUNDS FOR $B(k,n)$

Our upper bounds depend on the following conjecture.

Conjecture 4.1. *If $3 \le k \le n-1$, then*

$$2^{[\frac{n-1}{k}]}\frac{B(k,n+1)}{P(k,n+1)} < 2^{[\frac{n}{k}]}\frac{B(k,n)}{P(k,n)}. \qquad (4\text{–}1)$$

If $n > 10$, then we can improve (4–1) by adding a factor $\frac{1}{\sqrt{2}}$ on the right:

$$2^{[\frac{n-1}{k}]}\frac{B(k,n+1)}{P(k,n+1)} < 2^{[\frac{n}{k}]-\frac{1}{2}}\frac{B(k,n)}{P(k,n)}. \qquad (4\text{–}2)$$

Inequality (4–1) is weaker than Conjecture 3.2, but inequality (4–2) is stronger. We remark that it is not true that $B(k,n)/P(k,n)$ decreases as n increases, though (4–1) shows that this is true except perhaps when $n \equiv 0 \pmod{k}$. For example, we see from Table 3 that

$$\frac{B(3,7)}{P(3,7)} = 0.29 > \frac{B(3,6)}{P(3,6)} = 0.20.$$

Theorem 4.2. (Assume Conjectures 3.2, 4.1). *For* $3 \leq k \leq n-2$, *we have*

$$B(k,n) < 2^{[\frac{n-1}{k}]} P(k,n)/\sqrt{\pi 2^{n-1}}. \qquad (4\text{–}3)$$

Proof: It follows by induction from the case $k = 3$ of (3–1) in Conjecture 3.2 that

$$2^{[\frac{n-1}{3}]} \frac{B(k,n)}{P(k,n)} \leq 2^{[\frac{n-1}{k}]} \frac{B(3,n)}{P(3,n)} \quad \text{for } k \geq 4. \qquad (4\text{–}4)$$

If (4–3) is true for $k = 3$, then (4–4) gives

$$\frac{B(k,n)}{P(k,n)} \leq 2^{[\frac{n-1}{k}]-[\frac{n-1}{3}]} \frac{B(3,n)}{P(3,n)} < \frac{2^{[\frac{n-1}{k}]}}{\sqrt{\pi 2^{n-1}}},$$

so (4–3) holds for any $k \geq 4$. Thus it suffices to prove (4–3) for $k = 3$ and $n \geq 5$.

The values in Table 3 show that (4–3) is true for $k = 3$ and $5 \leq n \leq 10$, so we may assume $n > 10$ and use the stronger version (4–2) of Conjecture 4.1. We assume (4–3) for $k = 3$ and that $n > 10$. Then from (4–2) we get

$$\frac{B(3,n+1)}{P(3,n+1)} < 2^{[\frac{n}{3}]-[\frac{n-1}{3}]-\frac{1}{2}} \frac{B(3,n)}{P(3,n)} < \frac{2^{[\frac{n}{3}]}}{\sqrt{\pi 2^{n}}},$$

which completes the proof by induction. □

We can compare the lower bound of Theorem 3.3 with the upper bound of Theorem 4.2 using the estimate (easily derived from Stirling's formula for $k!$)

$$\binom{2^n}{2^{n-1}} = \frac{2^{2^n}}{\sqrt{\pi 2^{n-1}}} \left(1 - c2^{-n}\right),$$

where $c > 0$ is a constant.

The paper [Braeken et al. 05] is relevant to our work; the numerical results there concerning the cosets of the first-order Reed–Muller code are all consistent with our conjectures.

REFERENCES

[Braeken et al. 05] A. Braeken, Y. Borissov, S. Nikova, and B. Preneel. "Classification of Boolean Functions of 6 Variables or Less with Respect to Some Cryptographic Properties." In *Proceedings of ICALP 2005*, pp. 324–334, Springer Lecture Notes in Computer Science 3580. Berlin: Springer, 2005. An earlier version with more tables is available online (http://eprint.iacr.org/2004/248).

[Camion 79] P. Camion. "Etude de codes binaires abeliens modulaires autoduaux de petites longueurs." Research Report 350, Institut de Recherche d'Informatique et d'Automatique, 1979.

[Hou 94] X. Hou. "Classification of $R(3,8)/R(2,8)$." Manuscript, 1994.

[Hou 96] X. Hou. "$GL(m,2)$ Acting on $R(r,m)/R(r-1,m)$." *Discrete Mathematics* 149 (1996), 99–122.

[Kasami et al. 95] T. Kasami, T. Fujiwara, and Y. Desaki. "The Weight Distributions of Cosets of the Second-Order Reed–Muller Code of Length 128 in Third-Order Reed–Muller Code of Length 128." *IEICE Trans. Fundamentals* E79-A (1996), 600–608.

[Koumoto 04] T. Koumoto. "The Weight Distribution of BCH and Reed-Muller Codes." Available online (http://www.infsys.cne.okayama-u.ac.jp/~koumoto/wd/), 2004.

[Langevin 03] P. Langevin. "Classification of Boolean Cubic Forms." Available online (http://langevin.univ-tln.fr/cubics/), 2003.

[MacWilliams and Sloane 78] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes.* Amsterdam: North-Holland, 1978.

[McEliece 72] R. J. McEliece. "Weight Congruences for p-ary Cyclic Codes." *Discrete Math.* 3 (1972), 177–192.

[Sugino et al. 71] M. Sugino, Y. Ienaga, N. Tokura, and T. Kasami. "Weight Distribution of $(128, 64)$ Reed–Muller Code." *IEEE Trans. Inform. Theory* 17 (1971), 627–628.

[Sugita et al. 96] T. Sugita, T. Kasami, and T. Fujiwara. "Weight Distributions of the Third and Fifth Order Reed–Muller Codes of Length 512." Technical Report NAIST-IS-TR96006, Nara Institute of Science and Technology, 1996. Available online (http://isw3.aist-nara.ac.jp/IS/TechReport2/report/96006.ps).

T. W. Cusick, Department of Mathematics, State University of New York at Buffalo, Buffalo, NY 14260-2900
(cusick@buffalo.edu)

Y. Cheon, Department of Mathematics, Korea 3rd Military Academy, GoGyeong, YeongCheon, Korea, 770-849
(yhcrypt@yahoo.co.kr)

Received April 16, 2005; accepted in revised form September 8, 2006.