Autocorrelation Coefficients and Correlation Immunity of Boolean Functions

Yuriy Tarannikov *, Peter Korolev **, and Anton Botev ***

Mech. & Math. Department
Moscow State University
119899 Moscow, Russia

Abstract. We apply autocorrelation and Walsh coefficients for the investigation of correlation immune and resilient Boolean functions. We prove new lower bound for the absolute indicator of resilient functions that improves significantly (for $m > (n-3)/2$) the bound of Zheng and Zhang [18] on this value. We prove new upper bound for the number of nonlinear variables in high resilient Boolean function. This result supersedes the previous record. We characterize all possible values of resiliency orders for quadratic functions and give a complete description of quadratic Boolean functions that achieve the upper bound on resiliency. We establish new necessary condition that connects the number of variables, the resiliency and the weight of an unbalanced nonconstant correlation immune function and prove that such functions do not exist for $m > 0.75n - 1.25$. For high orders of m this surprising fact supersedes the well-known Bierbrauer–Friedman bound [8], [1] and was not formulated before even as a conjecture. We improve the upper bound of Zheng and Zhang [18] for the nonlinearity of high order correlation immune unbalanced Boolean functions and establish that for high orders of resiliency the maximum possible nonlinearity for unbalanced correlation immune functions is smaller than for balanced.

Keywords: Boolean functions, stream ciphers, correlation immunity, resiliency, nonlinearity, balancedness, Walsh Transform, autocorrelation coefficients, global avalanche characteristics, bounds.

1 Introduction

Different types of ciphers use Boolean functions. So, LFSR based stream ciphers use Boolean functions as a nonlinear combiner or a nonlinear filter, block ciphers use Boolean functions in substitution boxes and so on. Boolean functions used in ciphers must satisfy some specific properties to resist different attacks. One of the most important desired properties of Boolean functions in LFSR

* e-mails: yutaran@mech.math.msu.su, taran@vertex.inria.msu.ru
** e-mail: peter-korolev@mtu-net.ru
*** e-mail: stony_m@mail.ru

based stream ciphers is *correlation immunity* introduced by Siegenthaler [13]. Another important properties are nonlinearity, algebraic degree and so on. For Boolean functions used in block ciphers the most important properties are nonlinearity and differential (or autocorrelation) characteristics (propagation degree, avalanche criterion, the absolute indicator and so on) based on the autocorrelation coefficients of Boolean functions. Note that in recent research differential characteristics are considered as important for stream ciphers too.

Correlation immunity (or resiliency) is the property important in cryptography not only in stream ciphers. This is an important property if we want that the knowledge of some specified number of input bits does not give a (statistical) information about the output bit. In this respect such functions are considered in [6], [3] and other works.

Many works (see for example [5]) demonstrate that correlation immunity and autocorrelation characteristics are in strong contradiction. Some of results in our paper confirm it. Nevertheless, it appears that autocorrelation coefficients of a Boolean function is a power tool for the investigation of correlation immunity and other properties even without a direct relation to differential characteristics. The results of our paper demonstrate it.

In Section 2 we give preliminary concepts and notions. In Section 3 we prove new lower bound $\Delta_f \geq \left(\frac{2m-n+3}{n+1}\right) 2^n$ for the absolute indicator of resilient functions that improves significantly (for $m > (n-3)/2$) the bound of Zheng and Zhang [18] on this value. In Section 4 we prove that the number of nonlinear variables in n-variable $(n-k)$-resilient Boolean function does not exceed $(k-1)2^{k-2}$. This result supersedes the previous record $n \leq (k-1)4^{k-2}$ of Tarannikov and Kirienko [16]. As a consequence we give the sufficient condition on m and n that the absolute indicator of n-variable m-resilient function is equal to the maximum possible value 2^n. In Section 5 we characterize all possible values of resiliency orders for quadratic functions, i. e. functions with algebraic degree 2 in each variable. In Section 6 we give a complete description of quadratic n-variable m-resilient Boolean functions that achieve the bound $m \leq \frac{n}{2} - 1$. In Section 7 we establish new necessary condition that connects m, n and the weight of an n-variable unbalanced nonconstant mth order correlation immune function and prove that such functions do not exist for $m > 0.75n - 1.25$. For high orders of m this surprising fact supersedes the well-known Bierbrauer–Friedman bound [8], [1] and was not formulated before even as a conjecture. In Section 8 we prove that for $m \geq \frac{1}{2}n + \frac{1}{2}\log_2 n + \frac{1}{2}\log_2\left(\frac{\pi}{2}e^{8/9}\right) - 1$, $n \geq 12$, the nonlinearity of an unbalanced mth order correlation immune function of n variables does not exceed $2^{n-1} - 2^{m+1}$, and for $m \geq \frac{1}{2}n + \frac{3}{2}\log_2 n + \log_2\left(\frac{1}{4} + \frac{1}{n}\right) + \frac{1}{2}\log_2\left(\frac{\pi}{2}e^{8/9}\right) - 2$, $n \geq 24$, this nonlinearity does not exceed $2^{n-1} - 2^{m+2}$. These facts improve significantly correspondent results of Zheng and Zhang [18] and demonstrate that for higher orders of resiliency the maximum possible nonlinearity for balanced functions is greater than for unbalanced.

Along all paper we apply actively autocorrelation and Walsh coefficients for the investigation of correlation immune and resilient Boolean functions. Our new results demonstrate the power of this approach.

2 Preliminary concepts and notions

We consider ${F_2}^n$, the vector space of n-tuples of elements from F_2. An n-variable Boolean function is a map from ${F_2}^n$ into F_2. The *weight* of a vector x is the number of ones in x and is denoted by $|x|$. We say that the vector x *precedes* to the vector y and denote it as $x \preceq y$ if $x_i \leq y_i$ for each $i = 1, 2, \ldots, n$. The *scalar product* of vectors x and u is defined as $< x, u >= \sum_{i=1}^{n} x_i u_i$.

The *weight* $wt(f)$ of a function f on ${F_2}^n$ is the number of vectors x on ${F_2}^n$ such that $f(x) = 1$. A function f is said to be *balanced* if $wt(f) = wt(f \oplus 1) = 2^{n-1}$. A *subfunction* of the Boolean function f is a function f' obtained by substituting some constants for some variables in f.

It is well known that a function f on ${F_2}^n$ can be uniquely represented by a polynomial on F_2 whose degree in each variable in each term is at most 1. Namely,

$$f(x_1, \ldots, x_n) = \bigoplus_{(a_1,\ldots,a_n)\in {F_2}^n} g(a_1, \ldots, a_n) x_1^{a_1} \ldots x_n^{a_n}$$

where g is also a function on ${F_2}^n$. This polynomial representation of f is called the *algebraic normal form* (briefly, ANF) of the function and each $x_1^{a_1} \ldots x_n^{a_n}$ is called a *term* in ANF of f. The *algebraic degree* of f, denoted by $\deg(f)$, is defined as the number of variables in the longest term of f. The *algebraic degree of variable* x_i in f, denoted by $\deg(f, x_i)$, is the number of variables in the longest term of f that contains x_i. If $\deg(f, x_i) = 1$, we say that f depends on x_i *linearly*. If $\deg(f, x_i) \neq 1$, we say that f depends on x_i *nonlinearly*. A term of length 1 is called a *linear* term. If $\deg(f) \leq 1$ then f is called an *affine* function. If f is an affine function and $f(0) = 0$ then f is called a *linear* function.

Definition 1. *We say that the Boolean function f is* quadratic *if an algebraic degree of each variable in f is* 2*, i. e. if* $\deg(f, x_i) = 2$ *for each* $i = 1, 2, \ldots, n$.

The *Hamming distance* $d(x_1, x_2)$ between two vectors x_1 and x_2 is the number of components where vectors x_1 and x_2 differ. For two Boolean functions f_1 and f_2 on ${F_2}^n$, we define the distance between f_1 and f_2 by $d(f_1, f_2) = \#\{x \in {F_2}^n | f_1(x) \neq f_2(x)\}$. It is easy to see that $d(f_1, f_2) = wt(f_1 \oplus f_2)$. The minimum distance between f and the set of all affine functions is called the *nonlinearity* of f and denoted by $nl(f)$.

Definition 2. *The Walsh Transform of a Boolean function f is an integer-valued function over ${F_2}^n$ that can be defined as*

$$W_f(u) = \sum_{x\in {F_2}^n} (-1)^{f(x)+<u,x>}.$$

Walsh coefficients satisfy *Parseval's equation* $\sum_{u\in {F_2}^n} W_f^2(u) = 2^{2n}$.

Lemma 1. *Let f be an arbitrary Boolean function on ${F_2}^n$. Then*

$$wt(f) = 2^{n-1} - \frac{1}{2}W_f(0).$$

It is well known that $nl(f) = 2^{n-1} - \frac{1}{2}\max_{u \in F_2^n} |W_f(u)|$.

A Boolean function f on ${F_2}^n$ is said to be *correlation immune of order* m, with $1 \leq m \leq n$, if the output of f and any m input variables are statistically independent. This concept was introduced by Siegenthaler [13]. In equivalent non-probabilistic formulation the Boolean function f is called correlation immune of order m if $wt(f') = wt(f)/2^m$ for any its subfunction f' of $n - m$ variables. A balanced mth order correlation immune function is called an m-*resilient* function. In other words the Boolean function f is called m-resilient if $wt(f') = 2^{n-m-1}$ for any its subfunction f' of $n - m$ variables. In [9] a characterization of correlation immune functions by means of Walsh coefficients is given:

Theorem 1. [9] *A Boolean function f on ${F_2}^n$ is correlation-immune of order m if and only if $W_f(u) = 0$ for all vectors $u \in {F_2}^n$ such that $1 \leq |u| \leq m$.*

Theorem 2. [12] *If f is an mth order correlation immune function on ${F_2}^n$, $m \leq n-1$, then $W_f(u) \equiv 0 \pmod{2^{m+1}}$. Moreover, if f is m-resilient, $m \leq n-2$, then $W_f(u) \equiv 0 \pmod{2^{m+2}}$.*

Definition 3. *Let f be a Boolean function on ${F_2}^n$. For each $u \in {F_2}^n$ the autocorrelation coefficient of the function f at the vector u is defined as*

$$\Delta_f(u) = \sum_{x \in {F_2}^n} (-1)^{f(x)+f(x+u)}.$$

Zhang and Zheng [17] proposed the idea of Global Avalanche Characteristics (GAC). One of important indicators of GAC is *the absolute indicator.*

Definition 4. *Let f be a Boolean function on ${F_2}^n$. The absolute indicator of f is defined as*

$$\Delta_f = \max_{x \in {F_2}^n \setminus \{0\}} |\Delta_f(x)|.$$

3 New lower bound for the absolute indicator of resilient functions

In this section we prove new lower bound for the absolute indicator of resilient functions. At first, we establish an important technical formula. Note that this formula can be deduced from the relation $W_f^2(x) = \sum_{u \in {F_2}^n} (-1)^{<x,u>} \Delta_f(u)$ given in [5] and [4] but we prefer to give a direct proof in the Appendix A.

Theorem 3.

$$\Delta_f(u) = -2^n + 2^{1-n} \sum_{\substack{x \in {F_2}^n \\ <x,u> \equiv 0 \pmod 2}} W_f^2(x).$$

We denote by e_i the vector of the length n that has an one in ith component and zeroes in all other components.

Lemma 2. *Let f be an m-resilient Boolean function on ${F_2}^n$. Then*

$$\Delta_f \geq \left(\frac{2m - n + 2}{n} \right) 2^n.$$

Proof. We form the matrix B with n column writing in rows of B each binary vector $u \in {F_2}^n$ exactly $W_f^2(u)$ times. By Parseval's equality the matrix B contains exactly 2^{2n} rows. By Xiao Guo-Zhen–Massey spectral characterization [9] each row of the matrix B contains at most $n - m - 1$ zeroes. It follows that the total number of zeroes in B is at most $(n-m-1)2^{2n}$. Therefore, there exists some ith column in B that contains at most $\frac{(n-m-1)2^{2n}}{n}$ zeroes. By construction it follows that $\sum\limits_{\substack{x \in {F_2}^n \\ x_i = 0}} W_f^2(x) \leq \frac{(n-m-1)2^{2n}}{n}$. Then by Theorem 3 we have

$$\Delta_f(e_i) = -2^n + 2^{1-n} \sum_{\substack{x \in {F_2}^n \\ x_i = 0}} W_f^2(x) \leq -2^n + \frac{(n-m-1)}{n} 2^{n+1} \leq \frac{(n-2m-2)}{n} 2^n.$$

It follows that $\Delta_f \geq \left(\frac{2m-n+2}{n}\right) 2^n$. □

In the next theorem we improve the lower bound of Lemma 2.

Theorem 4. *Let f be an m-resilient Boolean function on ${F_2}^n$. Then $\Delta_f \geq \left(\frac{2m-n+3}{n+1}\right) 2^n$.*

Proof. Suppose that in the proof of Lemma 2 the matrix B contains exactly $h2^{2n}$ rows with less than $n - m - 1$ zeroes. Then repeating the arguments from the proof of Lemma 2 we have

$$\Delta_f \geq \left(\frac{2m - n + 2 + 2h}{n} \right) 2^n. \tag{1}$$

At the same time it is not hard to see that

$$\Delta_f(1 \ldots 1) = -2^n + 2^{1-n} \sum_{\substack{x \in {F_2}^n \\ |x| \equiv 0 \pmod 2}} W_f^2(x)$$

and

$$\Delta_f \geq |\Delta_f(1 \ldots 1)| \geq (1 - 2h)2^n. \tag{2}$$

The right part in (1) is increasing on h whereas the right part in (2) is decreasing on h. The right parts in (1) and (2) are equal when $h = \frac{n-m-1}{n+1}$. Therefore, $\Delta_f \geq \left(\frac{2m-n+3}{n+1}\right) 2^n$. □

In [19] Zheng and Zhang proved that for balanced mth order correlation immune function f on ${F_2}^n$ the bound $\Delta_f \geq \frac{2^n}{2^{n-m}-1}$ holds. It follows that $\Delta_f \geq 2^m + 2$. Our Theorem 4 improves significantly this result for $m > (n-3)/2$.

4 Upper bound for the number of nonlinear variables in high order resilient functions

In this section we prove the new upper bound for the number of nonlinear variables in high order resilient functions.

The next lemma is well-known.

Lemma 3. [11] *Let f be a Boolean function on ${F_2}^n$, $\deg(f) \geq 1$. Then*

$$2^{n-\deg(f)} \leq wt(f) \leq 2^n - 2^{n-\deg(f)}.$$

The next lemma is obvious.

Lemma 4. *Let f be a Boolean function on ${F_2}^n$, $\deg(f) \geq 1$. Then* $\deg(f(x) \oplus f(x+e_i)) \leq \deg(f(x)) - 1$.

Lemma 5. *Let f be a Boolean function on ${F_2}^n$, $\deg(f, x_i) \geq 2$. Then*

$$\sum_{\substack{u \in {F_2}^n \\ u_i = 0}} W_f^2(u) \geq 2^{2n-\deg(f)+1}.$$

Proof. By Theorem 3 using Lemmas 3 and 4 we have

$$-2^n + 2^{1-n} \sum_{\substack{u \in {F_2}^n \\ u_i = 0}} W_f^2(u) = \Delta_f(e_i) = \sum_{x \in {F_2}^n} (-1)^{f(x)+f(x+e_i)} =$$

$$2^n - 2wt(f(x) \oplus f(x+e_i)) \geq 2^n - 2\left(2^n - 2^{n-(\deg(f)-1)}\right) = -2^n + 2^{n-\deg(f)+2}.$$

It follows that $\sum_{\substack{u \in {F_2}^n \\ u_i = 0}} W_f^2(u) \geq 2^{2n-\deg(f)+1}$. □

Theorem 5. *Let f be an $(m = n-k)$-resilient Boolean function on ${F_2}^n$, $k \geq 2$, and $\deg(f, x_i) \geq 2$ for each $i = 1, \ldots, n$. Then $n \leq (k-1)2^{\deg(f)-1}$.*

Proof. We form the matrix B with n column writing in rows of B each binary vector $u \in {F_2}^n$ exactly $W_f^2(u)$ times. By Parseval's equality the matrix B contains exactly 2^{2n} rows. By Xiao Guo-Zhen–Massey spectral characterization [9] each row of the matrix B contains at most $k-1$ zeroes. It follows that the total number of zeroes in B is at most $(k-1)2^{2n}$. By Lemma 5 each column of B contains at least $2^{2n-\deg(f)+1}$ zeroes. Therefore $n \leq \frac{(k-1)2^{2n}}{2^{2n-\deg(f)+1}} = (k-1)2^{\deg(f)-1}$. □

Theorem 6. *Let f be an $(m = n-k)$-resilient Boolean function on ${F_2}^n$, $k \geq 2$, and $\deg(f, x_i) \geq 2$ for each $i = 1, \ldots, n$. Then $n \leq (k-1)2^{k-2}$.*

Proof. By Siegenthaler's Inequality [13] we have $\deg(f) \leq k-1$. This fact together with Theorem 5 follow the result. □

In [16] it is proved that $n \leq (k-1)4^{k-2}$. Our Theorem 6 improves significantly this result. Note that there exists $(n-k)$-resilient function on ${F_2}^n$, $n = 3 \cdot 2^{k-2} - 2$, that depends nonlinearly on all its n variables (see constructions in [14]).

Corollary 1. *Let f be an m-resilient Boolean function on ${F_2}^n$. If $n \geq (n-m-1)2^{n-m-2}$ then $\Delta_f = 2^n$.*

Proof. If $n > (n-m-1)2^{n-m-2}$ then by Theorem 6 the function f depends on some variable linearly, hence, $\Delta_f = 2^n$. If $n = (n-m-1)2^{n-m-2}$ and f depends on all its variables nonlinearly then according to the proofs of Theorems 5 and 6 we have that each row of the matrix B contains exactly $n-m-1$ zeroes. But in this case $|\Delta_f(1 \ldots 1)| = 2^n$, so, $\Delta_f = 2^n$. □

5 Resiliency orders of quadratic functions

In the next two sections we apply the autocorrelation coefficients for the analysis of quadratic Boolean functions, i. e. functions with algebraic degree 2 in each variable.

Lemma 6. *For any Boolean function g on ${F_2}^{n-1}$ the function $f(x_1, x_2, x_3, \ldots, x_n) = g(x_1 \oplus x_2, x_3, \ldots, x_n) \oplus x_1$ is balanced.*

Proof. We combine all vector from ${F_2}^n$ into pairs (y', y'') such that y' and y'' differ only in first and second components and coincide in all other components. Then $f(y') = f(y'') \oplus 1$ and $wt(f) = \sum\limits_{(y', y'')} (f(y') + f(y'')) = 2^{n-1}$. □

Lemma 7. *For each function $g(y_1, \ldots, y_n)$ on ${F_2}^n$ the function $f(x_1, \ldots, x_{2n}) = g(x_1 \oplus x_{n+1}, x_2 \oplus x_{n+2}, \ldots, x_n \oplus x_{2n}) \oplus x_1 \oplus x_2 \oplus \ldots \oplus x_n$ is $(n-1)$-resilient.*

Proof. Consider an arbitrary subfunction f' obtained from f by substitution of $n-1$ constants for some $n-1$ variables. Then there exists j such that both variables x_j and x_{n+j} remain free. Then f' has the form $f' = g'(\ldots, x_j \oplus x_{n+j}, \ldots) \oplus x_j$ and by Lemma 6 the function f is balanced. Hence, f is $(n-1)$-resilient. □

Theorem 7. *Quadratic m-resilient functions of n variables exist if and only if $m \leq \frac{n}{2} - 1$.*

Proof. Substitutuing to Theorem 5 the value $\deg(f) = 2$ we have $n \leq 2(n-m-1)$. It follows that $m \leq \frac{n}{2} - 1$. Now suppose that $m \leq \frac{n}{2} - 1$. Consider the function $f(x_1, \ldots, x_{2(n-m-1)}) = g(x_1 \oplus x_{n-m}, x_2 \oplus x_{n-m+1}, \ldots, x_{n-m-1} \oplus x_{2(n-m-1)}) \oplus x_1 \oplus x_2 \oplus \ldots \oplus x_{n-m-1}$ where g is some quadratic function. By Lemma 7 the function f is a $(2(n-m-1))$-variable $(n-m-2)$-resilient quadratic function. It is easy to check that if we substitute some constants for the variables $x_{n+1}, \ldots, x_{2(n-m-1)}$ in f then we obtain a desired n-variable m-resilient function. □

6 Complete description of quadratic Boolean functions with maximum resiliency order

In this section we give a complete description of quadratic resilient Boolean functions that achieve the bound $m \le \frac{n}{2} - 1$. It is obvious that for such functions n is even. Therefore in this section we consider for convenience $(N = 2n)$-variable $(m = n - 1)$-resilient functions.

Definition 5. *The notation* $f(x_1, \ldots, x_n) \stackrel{\sigma}{=} g(x_1, \ldots, x_n)$ *means that the Boolean functions* f *and* g *are equal up to permutation of indices of variables.*

Theorem 8. *Let* f *be an* $(N = 2n)$*-variable* $(m = n - 1)$*-resilient quadratic function. Then* $W_f(u) \neq 0$ *only if* $|u| = n$.

Proof. By Theorem 1 we have that $W_f(u) \neq 0$ only if $|u| \ge m + 1 = n$.

We form the matrix B with N columns writing in rows of B each binary vector $u \in {F_2}^N$ exactly $W_f^2(u)$ times. By Parseval's equality B contains exactly $2^{2N} = 2^{4n}$ rows. Each row has at most n zeroes, therefore the matrix B contains at most $n2^{4n}$ zeroes.

On the other hand, by Lemma 5 each column of the matrix B contains at least $2^{2N-\deg(f)+1} = 2^{4n-1}$ zeroes, i. e. the matrix B contains at least $2n2^{4n-1} = n2^{4n}$ zeroes.

Thus the matrix B contains exactly $n2^{4n}$ zeroes and each row of B has exactly n zeroes and n ones. □

Lemma 8. *Let* $e_{pq} = (0, \ldots, 0, \underset{p}{1}, 0, \ldots, 0, \underset{q}{1}, 0, \ldots, 0) \in {F_2}^n$, $p \neq q$, *and* f *is a quadratic function on* ${F_2}^n$. *Then* $\Delta_f(e_{pq}) \in \{0, \pm 2^n\}$ *and the next statements hold:*

$$\Delta_f(e_{pq}) = 2^n \iff f(x) = g(\ldots, x_p \oplus x_q, \ldots),\ g \text{ is quadratic},$$
$$\Delta_f(e_{pq}) = -2^n \iff f(x) = g(\ldots, x_p \oplus x_q, \ldots) \oplus x_p,\ g \text{ is quadratic}.$$

Proof.

We write the function f in the form $f(x) = \bigoplus\limits_{1 \le i < j \le n} a_{ij} x_i x_j \oplus \bigoplus\limits_{1 \le i \le n} b_i x_i \oplus c$ where $a_{ij} = a_{ji}$ and $a_{ii} = 0$.

Then $\Delta_f(e_{pq}) = \sum\limits_{x \in {F_2}^n} (-1)^{\bigoplus\limits_{i \neq p,q} (a_{pi} \oplus a_{qi}) x_i \oplus a_{pq}(x_p \oplus x_q \oplus 1) \oplus b_p \oplus b_q}$.

If the expression $\bigoplus\limits_{i \neq p,q} (a_{pi} \oplus a_{qi}) x_i \oplus a_{pq}(x_p \oplus x_q \oplus 1) \oplus b_p \oplus b_q$ contains at least one linear term x_k then we have $\Delta_f(e_{pq}) = 0$. If this expression does not contain linear terms, it means that $a_{pi} = a_{qi}$ for all i. Then $\Delta_f(e_{pq}) = 2^n(-1)^{b_p \oplus b_q}$ and the function f can be represented in the form

$$f(x) = \bigoplus_{\substack{i<j \\ i,j \neq p,q}} a_{ij} x_i x_j \oplus \bigoplus_{i \neq p,q} b_i x_i \oplus c \oplus \left(\bigoplus_{i \neq p,q} a_{pi} x_i \oplus b_q \right)(x_p \oplus x_q) \oplus (b_p \oplus b_q) x_p,$$

that completes the proof. □

Theorem 9. *Let $f(x_1,\dots,x_{2n})$ be a quadratic function on ${F_2}^{2n}$. Then*

$$\sum_{1\le p<q\le 2n} \Delta_f(e_{pq}) = -n2^{2n} \text{ if and only if}$$

$$f \stackrel{\sigma}{=} g(x_1\oplus x_{n+1},\dots,x_n\oplus x_{2n})\oplus x_1\oplus\ldots\oplus x_n$$

where $g(y_1,\dots,y_n)$ is a quadratic function on ${F_2}^n$.

Proof. Consider an arbitrary quadratic function f on ${F_2}^{2n}$. At the set of vertices $V=\{1,\dots,2n\}$ we construct the graph $G=(V,E)$ by the next rule: $(p,q)\in E$ if and only if $\Delta_f(e_{pq})\neq 0$.

Each connected component $H^t=(V^t,E^t)$ of this graph is a complete graph since by Lemma 8 we have that $(p,q)\in E^t$ if and only if $a_{pi}=a_{qi}$ for all i.

We divide V^t into two subsets $V_0^t \sqcup V_1^t$ such that $i\in V_{b_i}^t$. Let us denote $v_0^t := |V_0^t|,\ v_1^t := |V_1^t|$.

Then for p and q from the same subset of V^t by Lemma 8 we have $\Delta_f(e_{pq}) = 2^{2n}$ and for p and q from different subsets we have $\Delta_f(e_{pq}) = -2^{2n}$.

Let us estimate the sum

$$\sum_{(p,q)\in E^t}\Delta_f(e_{pq}) = 2^{2n}\left(\frac{v_0^t(v_0^t-1)}{2}+\frac{v_1^t(v_1^t-1)}{2}\right) - 2^{2n}v_0^tv_1^t =$$
$$= 2^{2n-1}((v_0^t-v_1^t)^2-(v_0^t+v_1^t)) \ge -2^{2n-1}|V^t|.$$

The equality is achieved only for $v_0^t = v_1^t = v^t$.

Hence, $\sum\limits_{1\le p<q\le 2n}\Delta_f(e_{pq}) \ge -2^{2n-1}\sum\limits_t |V^t| = -n2^{2n}$, moreover, the equality is achieved only if $v_0^t = v_1^t$ for all t.

Thus, if we have the equality $\sum\Delta_f(e_{pq}) = -n2^{2n}$ then it is possible to divide the set of all variables into the pairs (i_k^t, j_k^t) where $i_k^t\in V_0^t$, $j_k^t\in V_1^t$. Then the function will be represented in the form $f(x_1,\dots,x_{2n}) = g(x_{i_1^1}\oplus x_{j_1^1},\dots,x_{i_{v^1}^1}\oplus x_{j_{v^1}^1},\dots)\oplus x_{i_1^1}\oplus\ldots\oplus x_{i_{v^1}^1}\oplus\ldots$, i. e. in desired form.

Now suppose that the function has the form $g(x_1\oplus x_{n+1},\dots,x_n\oplus x_{2n})\oplus x_1\oplus\ldots\oplus x_n$. Then after the construction of the graph G and the partitioning it into components, we have $i\in V_{b_i}^t$ and $i+n\in V_{b_i\oplus 1}^t$ for all i, $i\le n$. It follows that $v_0^t = v_1^t$ for all t. □

Theorem 10. *Let $f(x_1,\dots,x_{2n})$ be an $(2n)$-variable $(n-1)$-resilient quadratic function. Then there exists a quadratic function $g(y_1,\dots,y_n)$ such that*

$$f(x_1,\dots,x_{2n}) \stackrel{\sigma}{=} g(x_1\oplus x_{n+1},\dots,x_n\oplus x_{2n})\oplus x_1\oplus\ldots\oplus x_n.$$

Proof. Substitute the equation from Theorem 3 into Theorem 9:

$$\sum_{1\le p<q\le 2n}\Delta_f(e_{pq}) = \sum_{p<q}\left(-2^{2n}+2^{1-2n}\sum_{\substack{x\in {F_2}^{2n}\\ <x,e_{pq}>\equiv 0 \pmod 2}} W_f^2(x)\right) =$$
$$= -n(2n-1)2^{2n} + 2^{1-2n}\sum_{p<q}\sum_{x_p=x_q} W_f^2(x).$$

By Theorem 8 for $|x| \neq n$ we have $W_f(x) = 0$, hence

$$\sum_{p<q} \sum_{x_p=x_q} W_f^2(x) = \sum_{p<q} \sum_{\substack{x_p=x_q \\ |x|=n}} W_f^2(x) = \sum_{|x|=n} \left(W_f^2(x) \sum_{p<q \,:\, x_p=x_q} 1 \right) =$$

$$= (n^2 - n) \sum_{|x|=n} W_f^2(x) = (n^2 - n)2^{4n}.$$

Therefore,

$$\sum_{1 \leq p<q \leq 2n} \Delta_f(e_{pq}) = -n(2n-1)2^{2n} + 2^{1-2n}(n^2 - n)2^{4n} = -n2^{2n}.$$

It follows by Theorem 9 that all $(2n)$-variable $(n-1)$-resilient quadratic functions have the given form. □

7 Nonexistence of unbalanced nonconstant mth order correlation immune Boolean functions on $F_2{}^n$ for $m > 0.75n - 1.25$

In this section we prove that unbalanced nonconstant mth order correlation immune Boolean functions on $F_2{}^n$ do not exist for $m > 0.75n - 1.25$. Similar statements are known for multioutputs functions (see [2], [10]) but for usual Boolean functions until now statements of such type were not formulated even as conjectures.

Theorem 11. *Let f be an arbitrary Boolean function on $F_2{}^n$. Let $w \in F_2{}^n \backslash \{0\}$. Then*

$$\sum_{\substack{x \in F_2{}^n \\ <x,w>=0}} W_f^2(x) = 2^{n-|w|} \sum_{\substack{u \in F_2{}^n \\ u \preceq w}} \Delta_f(u).$$

Proof. Summing $\Delta_f(u)$ over all u, $u \preceq w$, by Theorem 3 we have

$$\sum_{\substack{u \in F_2{}^n \\ u \preceq w}} \Delta_f(u) = \sum_{\substack{u \in F_2{}^n \\ u \preceq w}} \left(-2^n + 2^{1-n} \sum_{\substack{x \in F_2{}^n \\ <x,u> \equiv 0 \pmod 2}} W_f^2(x) \right) =$$

$$-2^{n+|w|} + 2^{1-n} \sum_{\substack{u \in F_2{}^n \\ u \preceq w}} \sum_{\substack{x \in F_2{}^n \\ <x,u> \equiv 0 \pmod 2}} W_f^2(x) =$$

$$-2^{n+|w|} + 2^{1-n} \left(2^{|w|} \sum_{\substack{x \in F_2{}^n \\ <x,w>=0}} W_f^2(x) + 2^{|w|-1} \sum_{\substack{x \in F_2{}^n \\ <x,w>>0}} W_f^2(x) \right) =$$

$$-2^{n+|w|} + 2^{1-n} \left(2^{|w|-1} \cdot 2^{2n} + 2^{|w|-1} \sum_{\substack{x \in F_2{}^n \\ <x,w>=0}} W_f^2(x) \right) = 2^{|w|-n} \sum_{\substack{x \in F_2{}^n \\ <x,w>=0}} W_f^2(x).$$

□

Theorem 12. *Let f be an arbitrary Boolean function on ${F_2}^n$. Then*

$$\sum_{\substack{u\in {F_2}^n\\ u\preceq w}} \Delta_f(u) = \sum_{f'} \left(2^{|w|} - 2wt(f')\right)^2$$

where the last sum is taken over all $2^{n-|w|}$ subfunctions f' of $|w|$ variables obtained from f by substituting constants for all x_i such that $w_i = 0$.

Proof.

$$\sum_{\substack{u\in {F_2}^n\\ u\preceq w}} \Delta_f(u) = \sum_{\substack{u\in {F_2}^n\\ u\preceq w}} \sum_{x\in {F_2}^n} (-1)^{f(x)+f(x+u)} =$$

$$\sum_{x\in {F_2}^n} \sum_{\substack{u\in {F_2}^n\\ u\preceq w}} (-1)^{f(x)+f(x+u)} = \sum_{f'} \sum_{x,y \text{ of } f'} (-1)^{f(x)+f(y)} =$$

$$\sum_{f'} \left(wt^2(f') + (2^{|w|} - wt(f'))^2 - 2wt(f')(2^{|w|} - wt(f')\right) = \sum_{f'} \left(2^{|w|} - 2wt(f')\right)^2.$$

□

Corollary 2. *Let f be an arbitrary Boolean function on ${F_2}^n$. Then*

$$\sum_{\substack{x\in {F_2}^n\\ <x,w>=0}} W_f^2(x) = 2^{n-|w|+2} \sum_{f'} \left(2^{|w|-1} - wt(f')\right)^2$$

where the last sum is taken over all $2^{n-|w|}$ subfunctions f' of $|w|$ variables obtained from f by substituting constants for all x_i such that $w_i = 0$.

Proof. It follows immediately from Theorems 11 and 12. □

Remark 1. If f is an $(n-k)$th order nonaffine correlation immune Boolean function on ${F_2}^n$ then by (2) we have $W_f(0) \equiv 0 \pmod{2^{n-k+1}}$. Therefore $W_f(0) \equiv 2^{n-i} \pmod{2^{n-i+1}}$ for some i, $i \in \{1, 2, \ldots, k-1\}$.

Theorem 13. *Let f be an unbalanced nonconstant $(n-k)$th order correlation immune Boolean function on ${F_2}^n$. Let $W_f(0) = \pm p \cdot 2^{n-i}$ where p is some odd positive integer, $i \in \{1, 2, \ldots, k-1\}$. Then*

$$\binom{n}{i} \le (2^{2i} - p^2) \binom{k-1}{i}. \tag{3}$$

Proof. By Lemma 1 we have that $\left|2^{n-1} - wt(f)\right| = p \cdot 2^{n-i-1}$. Let $w \in {F_2}^n$ be an arbitrary vector such that $|w| = i$. Then

$$\sum_{f'} \left|2^{i-1} - wt(f')\right| \ge \left|2^{n-1} - wt(f)\right| = p \cdot 2^{n-i-1}$$

where the sum is taken over all 2^{n-i} subfunctions f' of i variables obtained from f by substituting constants for all x_i such that $w_i = 0$. All terms in the sum are integer. It follows that

$$\sum_{f'} \left(2^{i-1} - wt(f')\right)^2 \geq \left(\left(\frac{p+1}{2}\right)^2 + \left(\frac{p-1}{2}\right)^2\right) \cdot 2^{n-i-1}.$$

Therefore by Corollary 2 we have

$$\sum_{\substack{x \in {F_2}^n \\ <x,w>=0}} W_f^2(x) \geq 2^{n-i+2} \cdot \left(\frac{p^2+1}{2}\right) \cdot 2^{n-i-1} = (p^2+1) \cdot 2^{2n-2i}.$$

Hence,

$$\sum_{\substack{x \in {F_2}^n \\ <x,w>=0}} W_f^2(x) - W_f^2(0) \geq 2^{2n-2i}. \tag{4}$$

Next, we form the matrix B with n columns writing in rows of B each binary vector $x \in {F_2}^n$ exactly $W_f^2(x)$ times. By Parseval's equality the matrix B contains exactly 2^{2n} rows. The total number of nonzero rows of B is $2^{2n} - p^2 \cdot 2^{2n-2i}$. By Xiao Guo-Zhen–Massey spectral characterization [9] each nonzero row of the matrix B contains at most $k-1$ zeroes. It follows that each nonzero row in B contains at most $\binom{k-1}{i}$ subsets of i zeroes. All nonzero rows in B contain at most $(2^{2n} - p^2 \cdot 2^{2n-2i}) \binom{k-1}{i}$ subsets of i zeroes. At the same time by (4) for any i columns in B there exist at least 2^{2n-2i} nonzero rows that contain only zeroes in these i columns. Therefore,

$$\frac{(2^{2n} - p^2 \cdot 2^{2n-2i}) \binom{k-1}{i}}{2^{2n-2i}} \geq \binom{n}{i}.$$

□

Corollary 3. *Let f be an mth order correlation immune Boolean function on ${F_2}^n$. Let $wt(f) = u \cdot 2^h$ where u is odd positive integer, h is integer. Then*

$$\binom{n}{h+1} \leq u(2^{n-h} - u) \binom{n-m-1}{h-m}.$$

Proof. It follows immediately from Theorem 13 and Lemma 1. □

Theorem 14. *Let f be an unbalanced nonconstant $(n-k)$th order correlation immune Boolean function on ${F_2}^n$. Then $n \leq 4k-5$.*

Proof. By Remark 1 we can assume that $W_f(0) \equiv 2^{n-i} \pmod{2^{n-i+1}}$ for some i, $i \in \{1, 2, \ldots, k-1\}$. Then by Theorem 13 we have

$$n(n-1)\ldots(n-i+1) \leq (2^{2i}-1)(k-1)(k-2)\ldots(k-i). \tag{5}$$

Suppose that $n \geq 4(k-1)$. Then $n(n-1)\ldots(n-i+1) \geq 2^{2i}(k-1)(k-2)\ldots(k-i)$ that contradicts to (5). □

Corollary 4. *For $m > 0.75n - 1.25$ there do not exist unbalanced nonconstant mth order correlation immune Boolean functions on ${F_2}^n$.*

It is easy to check that the 3-variable function f that takes the value 1 only at two vectors $(0,0,0)$ and $(1,1,1)$ is correlation immune of order 1. Therefore the bound in Corollary 4 is tight.

Remark 2. Until now Bierbrauer–Friedman bound [8], [1]

$$wt(f) \geq 2^n \frac{2(m+1)-n}{2(m+1)} \tag{6}$$

was the best known lower bound for the weight of high order correlation immune nonconstant functions. If we substitute $m > 0.75n-1.25$ to (6) we obtain $wt(f) > 2^n \frac{n-1}{3n-1}$. In fact, our Corollary 4 follows that in this case $wt(f) = 2^{n-1}$.

8 Tradeoff between correlation immunity and nonlinearity for unbalanced Boolean functions

In [12] Sarkar and Maitra proved (this result was obtained independently also in [14] and [18]) that for an n-variable mth order correlation immune Boolean function f, $n-m \geq 1$, the inequality $nl(f) \leq 2^{n-1} - 2^m$ holds. Moreover, if f is balanced (i. e. m-resilient), $n-m \geq 2$, then $nl(f) \leq 2^{n-1} - 2^{m+1}$. In [18] Zheng and Zhang proved that for unbalanced Boolean functions, $m \geq 0.6n - 0.4$, the nonlinearity $2^{n-1}-2^m$ can not be achieved. Therefore for an n-variable mth order correlation immune Boolean function f, $0.6n - 0.4 \leq m \leq n-1$, the inequality $nl(f) \leq 2^{n-1} - 2^{m+1}$ holds. (Note that by our Corollary 4 for $m > 0.75n - 1.25$ unbalanced n-variable mth order correlation immune functions do not exist at all!) At the same time in [15] Tarannikov gives the constructions of n-variable m-resilient Boolean functions with the nonlinearity $2^{n-1} - 2^{m+1}$ for $0.6n - 1 \leq m \leq n-2$. Thus, although the upper bound in [12] for unbalanced functions is higher than for balanced, nevertheless, at least for $0.6n - 0.4 \leq m \leq n-2$ the maximum possible nonlinearity of m-resilient Boolean functions is not less than the maximum possible nonlinearity of mth order correlation immune unbalanced Boolean functions. In this section we continue the investigations in this direction and give new improvements of upper bounds for the nonlinearity of high order correlation immune unbalanced Boolean functions. In our investigation we use the inequality (3) obtained in Theorem 13.

Theorem 15. *Let f be an unbalanced mth order correlation immune function on ${F_2}^n$. Suppose that $W_f(0) \equiv 2^{m+1} \pmod{2^{m+2}}$. Then for $n \geq 12$ the inequality*

$$m < \frac{1}{2}n + \frac{1}{2}\log_2 n + \text{const}$$

holds where $\text{const} = \frac{1}{2}\log_2\left(\frac{\pi}{2}e^{8/9}\right) - 1$.

The proof of Theorem 15 is given in the Appendix B.

Corollary 5. *Let f be an unbalanced mth order correlation immune function on ${F_2}^n$. If $m \geq \frac{1}{2}n+\frac{1}{2}\log_2 n+\frac{1}{2}\log_2\left(\frac{\pi}{2}e^{8/9}\right)-1$, $n \geq 12$, then $nl(f) \leq 2^{n-1}-2^{m+1}$.*

Proof. By Theorem 15 we have $W_f(0) \not\equiv 2^{m+1} \pmod{2^{m+2}}$. It follows that $|W_f(0)| \geq 2^{m+2}$. Therefore, $nl(f) = 2^{n-1}-\frac{1}{2}\max\limits_{x\in {F_2}^n}|W_f(x)| \leq 2^{n-1}-\frac{1}{2}|W_f(0)| \leq 2^{n-1}-2^{m+1}$. □

Theorem 16. *Let f be an unbalanced mth order correlation immune function on ${F_2}^n$. Suppose that $W_f(0) \equiv 2^{m+2} \pmod{2^{m+3}}$. Then for $n \geq 24$ the inequality*

$$m < \frac{1}{2}n + \frac{3}{2}\log_2 n + \log_2\left(\frac{1}{4}+\frac{1}{n}\right) + \text{const}$$

holds where $\text{const} = \frac{1}{2}\log_2\left(\frac{\pi}{2}e^{8/9}\right) - 2$.

The proof of Theorem 16 is given in the Appendix C.

Corollary 6. *Let f be an unbalanced mth order correlation immune function on ${F_2}^n$. If $m \geq \frac{1}{2}n + \frac{3}{2}\log_2 n + \log_2\left(\frac{1}{4}+\frac{1}{n}\right) + \frac{1}{2}\log_2\left(\frac{\pi}{2}e^{8/9}\right) - 2$, $n \geq 24$, then $nl(f) \leq 2^{n-1} - 2^{m+2}$.*

Proof. By Theorems 15 and 16 we have that $|W_f(0)| \geq 2^{m+3}$. Therefore, $nl(f) \leq 2^{n-1} - \frac{1}{2}|W_f(0)| \leq 2^{n-1} - 2^{m+2}$. □

Thus, we see that although the upper bounds in [12] for the nonlinearity of unbalanced functions is higher than for balanced, nevertheless, for higher m balanced functions are "better" than unbalanced in this respect.

The authors are grateful to Oktay Kasim-Zadeh for valuable advices on the analysis of inequality (7).

References

1. J. Bierbrauer, Bounds on orthogonal arrays and resilient functions, Journal of Combinatorial Designs, V. 3, 1995, pp. 179–183.
2. J. Bierbrauer, K. Gopalakrishnan, D. R. Stinson, Orthogonal arrays, resilient functions, error correcting codes and linear programming bounds, SIAM Journal of Discrete Mathematics, V. 9, 1996, pp. 424–452.
3. R. Canetti, Y. Dodis, S. Halevi, E. Kushilevitz, A. Sahai, Exposure-resilient functions and all-or-nothing transforms, In Advanced in Cryptology: Eurocrypt 2000, Proceedings, Lecture Notes in Computer Science, V. 1807, 2000, pp. 453–469.
4. A. Canteaut, C. Carlet, P. Charpin, C. Fontaine, Propagation characteristics and correlation-immunity of highly nonlinear Boolean functions, In Advanced in Cryptology: Eurocrypt 2000, Proceedings, Lecture Notes in Computer Science, V. 1807, 2000, pp. 507–522.
5. C. Carlet, Partially-bent functions, In Advanced in Cryptology: Crypto 1992, Proceedings, Lecture Notes in Computer Science, V. 740, 1992, pp. 280–291.
6. B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, R. Smolensky, The bit extraction problem or t-resilient functions, IEEE Symposium on Foundations of Computer Science, V. 26, 1985, pp. 396–407.

7. W. Feller, An introduction to probability theory and its applications, John Wiley & Sons, New York, 3rd edition, 1968.
8. J. Friedman, On the bit extraction problem, Proc. 33rd IEEE Symposium on Foundations of Computer Science, 1992, pp. 314–319.
9. Xiao Guo-Zhen, J. Massey, A spectral characterization of correlation-immune combining functions, IEEE Transactions on Information Theory, V. 34, No 3, May 1988, pp. 569–571.
10. V. Levenshtein, Split orthogonal arrays and maximum independent resilient systems of functions, Designs, Codes and Cryptography, V. 12, 1997, pp. 131–160.
11. F. J. Mac Williams, N. J. A. Sloane, The theory of error correcting codes, North-Holland, Amsterdam, 1977.
12. P. Sarkar, S. Maitra, Nonlinearity bounds and constructions of resilient Boolean functions, In Advanced in Cryptology: Crypto 2000, Proceedings, Lecture Notes in Computer Science, V. 1880, 2000, pp. 515–532.
13. T. Siegenthaler, Correlation-immunity of nonlinear combining functions for cryptographic applications, IEEE Transactions on Information theory, V. IT-30, No 5, 1984, p. 776–780.
14. Yu. Tarannikov, On resilient Boolean functions with maximal possible nonlinearity, Proceedings of Indocrypt 2000, Lecture Notes in Computer Science, V. 1977, pp. 19–30, Springer-Verlag, 2000.
15. Yu. Tarannikov, New constructions of resilient Boolean functions with maximal nonlinearity, Preproceedings of 8th Fast Software Encryption Workshop, Yokohama, Japan, April 2–4, 2001, pp.70-81.
16. Yu. Tarannikov, D. Kirienko, Spectral analysis of high order correlation immune functions, Proceedings of 2001 IEEE International Symposium on Information Theory ISIT2001, Washington, DC, USA, June 2001, p. 69, full version is available at Cryptology ePrint archive (http://eprint.iacr.org/), Report 2000/050, October 2000, 8 pp.
17. X. M. Zhang, Y. Zheng, GAC — the criterion for global avalanche characteristics and nonlinearity of cryptographic functions, Journal of Universal Computer Science, V. 1, 1995, pp. 136–150.
18. Y. Zheng, X. M. Zhang, Improved upper bound on the nonlinearity of high order correlation immune functions, Selected Areas in Cryptography, 7th Annual International Workshop, SAC2000, Lecture Notes in Computer Science, V. 2012, pp. 264–274, Springer-Verlag, 2001.
19. Y. Zheng, X. M. Zhang, New results on correlation immune functions, The 3nd International Conference on Information Security and Cryptology (ICISC 2000), Seoul, Korea, Lecture Notes in Computer Science, V. 2015, pp. 49–63, Springer-Verlag, 2001.

A Proof of Theorem 3

If $u = 0$ then obviously $\Delta_f(u) = 2^n$, and $\sum_{\substack{x \in {F_2}^n \\ <x,u> \equiv 0 \pmod 2}} W_f^2(x) = \sum_{x \in {F_2}^n} W_f^2(x) = 2^{2n}$, therefore, the equality holds. So, we can assume that $u \neq 0$. Next,

$$\sum_{\substack{x \in {F_2}^n \\ <x,u> \equiv 0 \pmod 2}} W_f^2(x) = \sum_{\substack{x \in {F_2}^n \\ <x,u> \equiv 0 \pmod 2}} \left(\sum_{y \in {F_2}^n} (-1)^{f(y)+<x,y>} \right)^2 =$$

$$\sum_{\substack{x\in {F_2}^n \\ <x,u>\equiv 0 \pmod 2}} \left(2^n + \sum_{y'\neq y''\in {F_2}^n} (-1)^{f(y')+f(y'')+<x,y'+y''>}\right) = 2^{2n-1} +$$

$$\sum_{y'\neq y''\in {F_2}^n} (-1)^{f(y')+f(y'')} \sum_{x\in {F_2}^n} \left(\frac{1}{2} + \frac{1}{2}(-1)^{<x,u>}\right)(-1)^{<x,y'+y''>} = 2^{2n-1} +$$

$$\frac{1}{2} \sum_{y'\neq y''\in {F_2}^n} (-1)^{f(y')+f(y'')} \left(\sum_{x\in {F_2}^n} (-1)^{<x,y'+y''>} + \sum_{x\in {F_2}^n} (-1)^{<x,u+y'+y''>}\right) =$$

$$2^{2n-1} + \frac{1}{2} \sum_{\substack{y',y''\in {F_2}^n \\ y'+y''=u}} (-1)^{f(y')+f(y'')} \left(0 + \sum_{x\in {F_2}^n} 1\right) =$$

$$2^{2n-1} + 2^{n-1} \sum_{y\in {F_2}^n} (-1)^{f(y)+f(y+u)} = 2^{2n-1} + 2^{n-1}\Delta_f(u).$$

□

B Proof of Theorem 15

For $i = k-1 = n-m-1$ in (3) we have

$$\binom{n}{i} < 4^i. \tag{7}$$

For each $i, 0 < i < n$, we have $\binom{n}{i} > (\frac{n}{i})^i$. It follows that if for some i the inequality (7) holds then the inequality $(\frac{n}{i})^i < 4^i$ holds too. Therefore $\frac{n}{i} < 4$ and $\frac{n}{4} < i$. Thus, we obtain the simplest bound on i: $i > \frac{n}{4}$.

By means of the lower and upper bounds for $n!$ (see[7])

$$\sqrt{2\pi}n^{n+1/2}e^{-n}e^{(12n+1)^{-1}} < n! < \sqrt{2\pi}n^{n+1/2}e^{-n}e^{(12n)^{-1}}$$

it is easy to deduce the inequality

$$\binom{n}{i} > \frac{2^{H(\frac{i}{n})n}}{\sqrt{2\pi n\frac{i}{n}(1-\frac{i}{n})}}\, e^{-\frac{1}{12n\frac{i}{n}(1-\frac{i}{n})}}, \tag{8}$$

that holds for any $0 < i < n$.

(Here $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the *entropy* of x, $0 < x < 1$).

If $\frac{n}{4} < i < \frac{n}{2}$ then $\frac{1}{4} < \frac{i}{n} < \frac{1}{2}$.

Consider the function $x(1-x)$. If $1/4 < x < 1/2$ then $3/16 < x(1-x) < 1/4$. It follows that for $\frac{n}{4} < i < \frac{n}{2}$ we have

$$\frac{3}{16} < \frac{i}{n}\left(1-\frac{i}{n}\right) < \frac{1}{4}.$$

Then

$$\frac{1}{\frac{i}{n}(1-\frac{i}{n})} > 4 \qquad \text{and} \qquad \frac{1}{\sqrt{\frac{i}{n}(1-\frac{i}{n})}} > 2,$$

i. e.

$$\frac{1}{\sqrt{2\pi\frac{i}{n}(1-\frac{i}{n})}} > \frac{2}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}. \tag{9}$$

Next,

$$\frac{1}{\frac{i}{n}(1-\frac{i}{n})} < \frac{16}{3}, \quad \text{it follows} \quad \frac{1}{12n\frac{i}{n}(1-\frac{i}{n})} < \frac{16}{12n\cdot 3} = \frac{4}{9n} \le \frac{4}{9},$$

since $n \ge 1$.

Therefore,

$$e^{-\frac{1}{12n\frac{i}{n}(1-\frac{i}{n})}} > e^{-\frac{4}{9}}. \tag{10}$$

From (8) using (9) and (10) we have for any i, $\frac{n}{4} < i < \frac{n}{2}$, that

$$\binom{n}{i} > \sqrt{\frac{2}{\pi}} e^{-\frac{4}{9}} \frac{2^{H(\frac{i}{n})n}}{\sqrt{n}}. \tag{11}$$

The inequalities (7) and (11) follow the inequality

$$4^i > \frac{\sqrt{\frac{2}{\pi}} e^{-\frac{4}{9}} 2^{H(\frac{i}{n})n}}{\sqrt{n}}. \tag{12}$$

Taking the logarithm in (12) we have

$$2i > H\left(\frac{i}{n}\right) n - \frac{1}{2}\log_2 n + \alpha$$

where $\alpha = \log_2\left(\sqrt{\frac{2}{\pi}} e^{-4/9}\right)$. Dividing by n we have

$$2\frac{i}{n} > H\left(\frac{i}{n}\right) - \frac{1}{2n}\log_2 n + \frac{\alpha}{n}.$$

Denoting $x = \frac{i}{n}$ we obtain the inequality

$$2x < H(x) - \frac{1}{2n}\log_2 n + \frac{\alpha}{n},$$

or

$$H(x) - 2x < a(n) \tag{13}$$

where $a(n) = \frac{1}{2n}\log_2 n - \frac{\alpha}{n}$.

Thus, the problem is reduced to the obtaining of lower bound for x satisfying (13) under the condition $1/4 < x < 1/2$.

Now put $y = \frac{1}{2} - x$. Then conditions on $x : 1/4 < x < 1/2$ transform into conditions on $y : 0 < y < 1/4$. To find the lower bound for x satisfying (13) is the same as to find the upper bound for y satisfying

$$H\left(\frac{1}{2} - y\right) - 2\left(\frac{1}{2} - y\right) < a(n),$$

or

$$H\left(\frac{1}{2} - y\right) - 1 + 2y < a(n). \tag{14}$$

By Taylor's formula

$$H\left(\frac{1}{2} - y\right) = H\left(\frac{1}{2}\right) - H'\left(\frac{1}{2}\right) y + \frac{1}{2} H''(\xi) y^2 \tag{15}$$

where ξ is some number from the interval $1/2 - y < \xi < 1/2$. Taking into account that $y < 1/4$ we have $1/4 < \xi < 1/2$.

We differentiate and find that $H'(x) = \log_2 \frac{1-x}{x}$, $H''(x) = -\frac{1}{\ln 2} \frac{1}{x(1-x)}$.

It follows $H'(\frac{1}{2}) = 0$, also for $1/4 < \xi < 1/2$ the inequality $H''(\frac{1}{4}) < H''(\xi) < H''(\frac{1}{2})$ holds, in particular, $H''(\xi) > -\frac{16}{3 \ln 2}$ (the function $\frac{-1}{x(1-x)}$ increases for $0 < x < 1/2$). Also we take into account that $H(\frac{1}{2}) = 1$.

From (15) we have for any y, $0 < y < 1/4$,

$$H\left(\frac{1}{2} - y\right) > H\left(\frac{1}{2}\right) - H'\left(\frac{1}{2}\right) y + H''\left(\frac{1}{4}\right) y^2 = 1 - \frac{8}{3 \ln 2} y^2.$$

Taking into account the last inequality in (14) we have

$$1 - \frac{8}{3 \ln 2} y^2 - 1 + 2y < a(n),$$

or

$$0 < \frac{8}{3 \ln 2} y^2 - 2y + a(n). \tag{16}$$

The inequality (16) is quadratic with respect to y and depends on the parameter n. The coefficient in quadratic term is positive, therefore y can be determined from the conditions $y < y_1$ or $y > y_2$ where $y_1 < y_2$ are roots of characteristic equation. The second condition is irrelevant and does not correspond to the sense of this problem. A discriminant is equal to

$$1 - \frac{8}{3 \ln 2} a(n).$$

Note that $\sqrt{\frac{2}{\pi}} e^{-4/9} < 1$, it follows $\alpha < 0$. Let $\beta = -\alpha > 0$. Then $a(n) = \frac{1}{2n} \log_2 n + \frac{\beta}{n}$ where $\beta > 0$.

Thus, it is sufficient to solve the inequality

$$0 < \gamma y^2 - 2y + b(n) \tag{17}$$

where $\gamma = \frac{8}{3\ln 2}$.

Positiveness of a discriminant means that $1 - \gamma b(n) > 0$ or $b(n) < 1/\gamma$, i. e.

$$\frac{1}{2n}\log_2 n + \frac{\beta}{n} < \frac{1}{\gamma}. \tag{18}$$

The function $\frac{\ln x}{x}$ has the maximum for $x = e$. Let $n \geq 12$, then $\frac{1}{2n}\log_2 n \leq \frac{1}{24}\log_2 12$. Hence, it is sufficient to demonstrate that

$$\frac{1}{24}\log_2 12 + \frac{\log_2(\sqrt{\frac{\pi}{2}}e^{4/9})}{12} < \frac{3\ln 2}{8}$$

or

$$\frac{1}{3}(2 + \log_2 3) + \frac{2}{3}(\log_2 \sqrt{\frac{\pi}{2}}e^{4/9}) < 3\ln 2. \tag{19}$$

The right part of (19) is greater than 2 since $e^2 < 8 = e^{3\ln 2}$. Consider the left part of (19). It is equal to

$$\frac{2}{3} + \frac{\log_2 3}{3} + \frac{2}{3}\left(\log_2 \sqrt{\frac{\pi}{2}}e^{4/9}\right) < \frac{2}{3} + \frac{1}{3}\left(\log_2 \frac{3\pi e}{2}\right).$$

The product $\pi e < 10$, therefore $\frac{3\pi e}{2} < 16$. Hence, the left part of (19) is less than 2. It follows that for $n \geq 12$ a discriminant of the equation (17) is positive and required upper bound for y follows from the inequality

$$y < y_1 = \frac{1 - \sqrt{1 - \frac{8}{3\ln 2}a(n)}}{\frac{8}{3\ln 2}} \leq \frac{1 - 1 + \frac{8}{3\ln 2}a(n)}{\frac{8}{3\ln 2}} = a(n)$$

where y_1 is a root of the equation correspondent to the inequality (16). Pointing in a view that $y = \frac{1}{2} - x = \frac{1}{2} - \frac{i}{n} = \frac{1}{2} - \frac{n-m-1}{n} = \frac{m+1}{n} - \frac{1}{2}$ we have:

$$\frac{m+1}{n} - \frac{1}{2} < \frac{1}{2n}\log_2 n + \frac{1}{2n}\log_2\left(\frac{\pi}{2}e^{8/9}\right).$$

For $n \geq 12$ it follows

$$m < \frac{1}{2}n + \frac{1}{2}\log_2 n + \text{const}$$

where $\text{const} = \frac{1}{2}\log_2(\frac{\pi}{2}e^{8/9}) - 1$. □

C Proof of Theorem 16

For $i = k - 2 = n - m - 2$ in (3) we have

$$\binom{n}{i} \leq (4^i - 1)(i + 1). \tag{20}$$

As in the previous proof we use the inequality (8), the bounds (9) and (10) valid for sufficiently high n and the inequality (11). Combining (11) and (20) we have

$$4^i(i+1) > \sqrt{\frac{2}{\pi}} e^{-\frac{4}{9}} \frac{2^{H(\frac{i}{n})n}}{\sqrt{n}}.$$

Taking the logarithm in the last inequality we have:

$$2i + \log_2(i+1) > H\left(\frac{i}{n}\right) n - \frac{1}{2}\log_2 n + \alpha,$$

$\alpha = \log_2\left(\sqrt{\frac{2}{\pi}} e^{-\frac{4}{9}}\right)$.

Introducing new variable $x = \frac{i}{n}$ and dividing by n we obtain

$$2x + \frac{\log_2(x + \frac{1}{n})}{n} > H(x) - \frac{3}{2}\frac{\log_2 n}{n} + \frac{\alpha}{n},$$

Taking into account that $\log_2(x + \frac{1}{n}) \geq \log_2(\frac{1}{4} + \frac{1}{n})$ we have:

$$H(x) - 2x < a(n)$$

where $a(n) = \frac{\log_2(\frac{1}{4}+\frac{1}{n})}{n} + \frac{3}{2}\frac{\log_2 n}{n} - \frac{\alpha}{n}$.

This inequality is analogous to the inequality (13); the only difference is in the function $a(n)$. Using the reasonings completely analogous to the reasoning in the previous proof we deduce the inequality

$$0 < \gamma y^2 - 2y + b(n) \tag{21}$$

where $\gamma = \frac{8}{3\ln 2}$, $y = \frac{1}{2} - x$, $b(n) = \frac{3}{2}\frac{\log_2 n}{n} + \frac{\log_2(\frac{1}{4}+\frac{1}{n})}{n} + \frac{\beta}{n}$, $\beta = -\alpha$.

The solutions of this inequality satisfy $y < a(n)$ (see Appendix B). It means that $\frac{1}{2} - \frac{n-(m+2)}{n} < \frac{\log_2(\frac{1}{4}+\frac{1}{n})}{n} + \frac{3}{2}\frac{\log_2 n}{n} - \frac{\alpha}{n}$ or rewriting

$$m < \frac{1}{2}n + \frac{3}{2}\log_2 n + \log_2\left(\frac{1}{4} + \frac{1}{n}\right) + c,$$

$c = -\alpha - 2$.

Now we need only to know for which n this inequality is satisfied, i. e. beginning with which n a discriminant of inequality (21) is nonnegative, or $b(n) < 1/\gamma$, or

$$\frac{3\log_2 n}{2n} + \frac{\log_2(\frac{1}{4} + \frac{1}{n})}{n} + \frac{\log_2\left(\sqrt{\frac{\pi}{2}} e^{4/9}\right)}{n} < \frac{3\ln 2}{8}.$$

Computer analysis shows that this inequality is true beginning with $n = 24$. It completes the proof. □