Algebraic Immunity for Cryptographically Significant Boolean Functions: Analysis and Construction

Claude Carlet, Deepak Kumar Dalai, Kishan Chand Gupta, and Subhamoy Maitra

***Abstract*—Recently, algebraic attacks have received a lot of attention in the cryptographic literature. It has been observed that a Boolean function f used as a cryptographic primitive, and interpreted as a multivariate polynomial over F_2, should not have low degree multiples obtained by multiplication with low degree nonzero functions. In this paper, we show that a Boolean function having low nonlinearity is (also) weak against algebraic attacks, and we extend this result to higher order nonlinearities. Next, we present enumeration results on linearly independent annihilators. We also study certain classes of highly nonlinear resilient Boolean functions for their algebraic immunity. We identify that functions having low-degree subfunctions are weak in terms of algebraic immunity, and we analyze some existing constructions from this viewpoint. Further, we present a construction method to generate Boolean functions on n variables with highest possible algebraic immunity $\lceil \frac{n}{2} \rceil$ (this construction, first presented at the 2005 Workshop on Fast Software Encryption (FSE 2005), has been the first one producing such functions). These functions are obtained through a doubly indexed recursive relation. We calculate their Hamming weights and deduce their nonlinearities; we show that they have very high algebraic degrees. We express them as the sums of two functions which can be obtained from simple symmetric functions by a transformation which can be implemented with an algorithm whose complexity is linear in the number of variables. We deduce a very fast way of computing the output to these functions, given their input.**

***Index Terms*—Algebraic attacks, annihilators, Boolean functions, nonlinearity, stream ciphers, Walsh spectrum.**

I. Introduction

A very well studied model of stream cipher is the nonlinear combiner model, where the outputs to several linear feedback shift registers (LFSRs) are combined using a nonlinear Boolean function to produce the key stream. This model has undergone a lot of cryptanalysis and to resist those attacks, different design criteria have been proposed for both the LFSRs and the combining Boolean function. The main criteria on the combining function are balancedness, a high algebraic degree, a high nonlinearity, and correlation immunity. Another model is the filter generator, in which the content of some of the flip-flops in a single (longer) LFSR constitute the input to a nonlinear Boolean function which produces the keystream. This model is theoretically equivalent to the combiner model, but the attacks do not work quite similarly on each system. The main criteria on the filtering function are balancedness, a high algebraic degree, and a high nonlinearity. There are a large number of papers studying all of these criteria and one may refer to [12], [34], [43], [14], [50], and the references in these papers for more details.

Very recently, a new attack that uses cleverly overdefined systems of multivariate nonlinear equations to recover the secret key has gained a lot of attention (the idea of using such systems comes from Shannon [45], but the improvement in the efficiency of the method is recent). It is known as algebraic attack [3], [4], [22]–[25], [35], [38]. Given a Boolean function f on n variables, different kinds of scenarios related to low-degree multiples of f have been studied in [24], [38]. The core of the analysis is to find minimum (or low) degree annihilators of f or of $1+f$, i.e., to find minimum (or low) degree nonzero functions g such that $f * g = 0$ or $(1+f) * g = 0$. To mount the algebraic attack, one needs only low-degree annihilators [24], [38] of $f, 1+f$ (at least one and, better, as many linearly independent ones as possible).

In this paper, we study the immunity of Boolean functions against algebraic attacks, called the algebraic immunity. We show some relationships between the algebraic immunity and the nonlinearity of a Boolean function by proving that a Boolean function with low nonlinearity must have low algebraic immunity. This result relates the algebraic immunity to the Walsh spectrum of a Boolean function. We also present enumeration results on the number of annihilators.

We study the algebraic immunity of those functions satisfying the criteria recalled above. We present experimental results on highly nonlinear resilient (that is, balanced and correlation immune) functions which are rotation symmetric [32], [47], [48], [33], [37]. The experiments have been done using Algorithm 1 [38] on functions of seven, eight, and nine variables and their complements. The results found are encouraging, which shows that there exist highly nonlinear resilient functions that are also good in terms of their algebraic immunity (see also [20]).

So far, little attempt has been made to provide construction of Boolean functions that can resist algebraic attacks. One attempt in this direction is to analyze some existing construction methods that can provide Boolean functions with some other cryptographic properties to see how good they are in terms of algebraic immunity [5]–[7], [16]. We study different construction methods of resilient functions: *primary constructions*, which produce functions directly, and *secondary constructions*,

Manuscript received June 29, 2005; revised February 9, 2006. The material in this paper was presented in part at Indocrypt, Chennai, India, December 2004, and at 2005 Workshop on Fast Software Encryption, Paris, France, February 2005. Some results from this paper appeared in an initial form in [27] and [28].

C. Carlet is with INRIA, Project CODES, BP 105-78153, Le Chesnay Cedex, France (e-mail: claude.carlet@inria.fr); he is also with the University of Paris 8 (MAATICAH).

D. K. Dalai and S. Maitra are with the Applied Statistics Unit, Indian Statistical Institute, Kolkata, Pin 700 108, India (e-mail: deepak_r@isical.ac.in; subho@isical.ac.in).

K. C. Gupta is with the Centre for Applied Cryptographic Research, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: kgupta@math.uwaterloo.ca).

Communicated by E. Okamoto, Associate Editor for Complexity and Cryptography.

which give new functions from previously designed ones. We have experimentally studied some functions which are of Maiorana–McFarland type [11], i.e., which can be seen as concatenations of affine functions. We also show that, if a Boolean function has low-degree subfunctions, then it is not good in terms of algebraic immunity. This completes the analysis on Maiorana–McFarland type functions presented in [38].

Analyzing existing construction methods to see how good they are in terms of algebraic immunity is only an *ad hoc* attempt, as these existing construction methods are not meant for getting good algebraic immunity. In this paper, we provide a construction method where the algebraic immunity is the main concern. We introduce the (primary) construction of a $2k$-variable Boolean function with algebraic immunity provably equal to k (that is, optimal). This construction, initially presented at the 2005 Workshop on Fast Software Encryption (FSE 2005) and presented there as a secondary one, has been originally the first one producing functions with optimum algebraic immunity. The construction is iterative in nature (a function with two more variables is constructed at each step). This function can then be used in a secondary construction, to obtain a balanced function with highest possible algebraic immunity or with a reasonably high algebraic immunity, nonlinearity, and (if necessary) resiliency order. We show that the function has very high algebraic degree. We also give an algorithm permitting to deduce the function from two symmetric functions, which allows to have a very fast way (whose complexity is linear in the number of variables) of computing the output to the function, given its input. This was necessary so that the function can be efficiently used. Indeed, to make the complexity of algebraic attacks greater than 2^{128} (i.e., more complex than exhaustive search) a strict minimum seems an algebraic immunity of 8—see Section II, Remark 1—which implies at least 15 variables for the function itself, plus the number of variables necessary for applying a secondary construction ensuring balancedness and good nonlinearity, and if necessary good resiliency; the efficiency of the stream cipher is then a real challenge.

Other fast computable functions exist with optimal algebraic immunity (they have been given originally in [29] and a little later in [10], with further examples). They are symmetric and present therefore a risk if attacks using this peculiarity can be found in the future. Our functions do not have this drawback.

As this current effort has been an ongoing work for some time, a lot of issues have been raised in this area in the meantime. One should first note that by algebraic immunity we mean the resistance against standard algebraic attacks, done in a particular way, i.e., using linearization. One does not need linearization if algorithms using Gröbner bases can be properly exploited. However, algebraic immunity is still a relevant notion, since cryptosystems must at least resist the attacks by linearization, and since the complexity of the attacks by Gröbner bases (which are faster than the attacks by linearization) is difficult to evaluate. Further, it should be noted that based on some recent works related to fast algebraic attacks [2], [25], [9], one should concentrate more carefully on the design parameters of Boolean functions for proper resistance. This is the reason why, in one of the recent papers [29], the term of "annihilator immunity" is used instead of "algebraic immunity." However, even in the case of fast algebraic attacks, the algebraic immunity plays an important role, as shown in [1].

II. Preliminaries

A Boolean function on n variables may be viewed as a mapping from F_2^n into F_2, the finite field with two elements. We denote by B_n the set of all n-variable Boolean functions. One of the standard representations of a Boolean function $f(x_1,\ldots,x_n)$ is by the output column of its *truth table*, i.e., a binary string of length 2^n

$$f = [f(0,0,\ldots,0), f(1,0,\ldots,0), f(0,1,\ldots,0), f(1,1,\ldots,0),\ldots, f(1,1,\ldots,1)].$$

The set of $x \in F_2^n$ for which $f(x)=1$ (respectively, $f(x)=0$) is called the on-set (respectively, offset), denoted by 1_f (respectively, 0_f). We say that a Boolean function f is balanced if the truth table contains an equal number of 1's and 0's.

The Hamming weight of a binary string S is the number of ones in the string. This number is denoted by $\mathrm{wt}(S)$. The Hamming distance between two strings, S_1 and S_2 is denoted by $d(S_1,S_2)$ and is the number of places where S_1 and S_2 differ. Note that $d(S_1,S_2)=\mathrm{wt}(S_1+S_2)$ (by abuse of notation, we also use $+$ to denote the addition in F_2, i.e., the XOR).

Any Boolean function has a unique representation as a multivariate polynomial over F_2, called the algebraic normal form (ANF)

$$f(x_1,\ldots,x_n) = a_0 + \sum_{1\le i\le n} a_i x_i + \sum_{1\le i<j\le n} a_{ij}x_ix_j + \cdots + a_{12\ldots n}x_1x_2\ldots x_n$$

where the coefficients $a_0, a_i, a_{ij},\ldots,a_{12\ldots n}$ belong to $\{0,1\}$. The algebraic degree $\deg(f)$ is the number of variables in the highest order term with nonzero coefficient. A Boolean function is affine if there exists no term of degree strictly greater than 1 in the ANF and the set of all affine functions is denoted by A_n. An affine function with constant term equal to zero is called a linear function.

It is known that a Boolean function should have high algebraic degree to be cryptographically secure [31]. Further, it has been identified recently that it should not have a low-degree multiple. More precisely, it is shown in [24] that, given any n-variable Boolean function f, it is always possible to get a Boolean function g with degree at most $\lceil \frac{n}{2} \rceil$ such that $f*g$ has degree at most $\lceil \frac{n}{2} \rceil$. Here the functions are considered to be multivariate polynomials over F_2 and $f*g$ is the polynomial multiplication over F_2. Thus, while choosing a function f, the cryptosystem designer should be careful that it should not happen that the degree of $f*g$ falls much below $\lceil \frac{n}{2} \rceil$ with a nonzero function g whose degree is also much below $\lceil \frac{n}{2} \rceil$. In fact, as observed in [24], [38], it is enough to check that f and $f+1$ do not admit nonzero annihilators of such low degrees.

Definition 1: Given $f \in B_n$, define

$$AN(f) = \{g \in B_n \mid f*g = 0\}.$$

Any function $g \in AN(f)$ is called an annihilator of f.

To check that a function has good algebraic immunity, it is necessary and sufficient to check that f and $f+1$ do not admit nonzero annihilators of low degrees. Indeed, if f or $f+1$ has an annihilator g of low degree d, then $f * g$ either is null or equals g and therefore has degree at most d; conversely, if we have $f * g = h$ where $g \neq 0$ and where g and h have degrees at most d, then either $g = h$, and then g is an annihilator of $f+1$, or $g \neq h$, and we have then $f * g = f * h$ by multiplying both terms of the equality $f * g = h$ by f, which proves that $f * (g+h) = 0$ and shows that $g+h$ is a nonzero annihilator of f of degree at most d.

Definition 2: Given $f \in B_n$, we define its algebraic immunity as the minimum degree of all nonzero annihilators of f or $f+1$, and we denote it by $\mathcal{AI}_n(f)$.

Note that $\mathcal{AI}_n(f) \leq \deg(f)$, since $f * (1+f) = 0$. As f or $1+f$ must have an annihilator at a degree $\leq \lceil \frac{n}{2} \rceil$ [24], we have $\mathcal{AI}_n(f) \leq \lceil \frac{n}{2} \rceil$.

Remark 1: Let an n-variable function f, with algebraic immunity $\lceil \frac{n}{2} \rceil$ be used as a filtering function on a linear automaton (e.g., an LFSR) with $m \geq 2k$ states, where k is the length of the key (otherwise, it is known that the system is not robust). Then the complexity of an algebraic attack using one annihilator of degree $\lceil \frac{n}{2} \rceil$ is roughly

$$7\left(\binom{m}{0}+\cdots+\binom{m}{\lceil \frac{n}{2} \rceil}\right)^{\log_2(7)} \approx 7\left(\binom{m}{0}+\cdots+\binom{m}{\lceil \frac{n}{2} \rceil}\right)^{2.8}$$

(see [24]). Let us choose $k = 128$ (which is usual) and $m = 256$, then the complexity of the algebraic attack is greater than the complexity of an exhaustive search, that is 2^{128}, for $n \geq 15$. If the attacker knows several linearly independent annihilators of degree $\lceil \frac{n}{2} \rceil$, then the number of variables must be enhanced.

Remark 2: There are some recent works [2], [25], [9], [1], based on which one may need to consider the situations further to annihilators. Consider, for instance, the situation when $f * h = 0$, and h is a lowest degree annihilator of f. Let the degree of h be d_h. Then generally we expect that the cryptanalysis will be performed considering the annihilator h and its degree is an important parameter in the complexity of the attack. Consider that one has designed a scheme considering this scenario. However, it may very well happen that $f * g = H$, where $\deg(H) = \deg(h)$, but $\deg(g) < \deg(h)$ and in such an event one may get a better attack (with lower complexity) using g. This has been exploited in [26] to present an attack on SFINKS [8].

In this work, we are concentrating on algebraic immunity as defined in Definition 2. One should note that algebraic immunity (as in Definition 2) is not a property that can resist all kinds of algebraic attacks, but clearly this is a necessary one. Our studies in this paper are based in the scope of this definition and we leave it as open problem to see how these analyses can be extended keeping in mind the properties emerged to resist fast algebraic attacks.

The *nonlinearity* of an n-variable function f is its distance from the set of all n-variable affine functions, i.e.,

$$nl(f) = \min_{g \in A_n} (d(f,g)).$$

Boolean functions used in cryptographic systems must have high nonlinearity to withstand linear and correlation attacks [31], [12].

It is known that there are highly nonlinear Boolean functions of low degrees; as example, there exist quadratic bent functions that have degree 2 and maximum possible nonlinearity $2^{n-1} - 2^{\frac{n}{2}-1}$, when n is even. Such functions f, as they are by themselves of low algebraic degree, will have low values of algebraic immunity $\mathcal{AI}_n(f)$. On the other hand, we may have Boolean functions of low nonlinearity with high algebraic degree. Interestingly, if we replace the algebraic degree by the algebraic immunity, the situation changes. In this paper, we show that, if a function has low nonlinearity, then it must have a low value of $\mathcal{AI}_n(f)$. This implies that if one chooses a function with good value of $\mathcal{AI}_n(f)$, this will automatically provide a nonlinearity which is not low. However, it does not assure that the nonlinearity is very high (see Section III). Hence, the algebraic immunity property takes care of two fundamental properties of a Boolean function, algebraic degree and nonlinearity, but it does this incompletely in the case of nonlinearity. We will recall also that this property stays unchanged with respect to linear transformation unlike correlation immunity or propagation characteristics.

Many properties of Boolean functions can be described by the Walsh transform. Let $x = (x_1, \ldots, x_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$ both belonging to F_2^n and $x \cdot \omega = x_1\omega_1 + \cdots + x_n\omega_n$. Let $f(x)$ be a Boolean function on n variables. Then the *Walsh transform* of $f(x)$ is an integer valued function over F_2^n which is defined as

$$W_f(\omega) = \sum_{x \in F_2^n} (-1)^{f(x)+x\cdot\omega}.$$

A Boolean function f is balanced if and only if $W_f(0) = 0$. The nonlinearity of f is given by

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{\omega \in F_2^n} |W_f(\omega)|.$$

Correlation-immune functions and resilient functions are two important classes of Boolean functions. A function is m-resilient (respectively, mth-order correlation immune) if and only if its Walsh transform satisfies $W_f(\omega) = 0$, for $0 \leq \mathrm{wt}(\omega) \leq m$ (respectively, $1 \leq \mathrm{wt}(\omega) \leq m$).

Following the notation as in [42], [43], [48] we use (n, m, d, σ) to denote n-variable, m-resilient function with degree d and nonlinearity σ. Further, by $[n, m, d, \sigma]$ we denote unbalanced n-variable, m-th order correlation immune function with degree d and nonlinearity σ.

III. Algebraic Immunity and Walsh Spectrum

Toward proving the results relating algebraic immunity and the nonlinearities of a Boolean function, we first present the following result, where we relate the algebraic degree with the weight of the function.

Theorem 1: Let $f \in B_n$ and $\mathcal{AI}_n(f) > d$. Then

$$\sum_{i=0}^{d} \binom{n}{i} \leq \mathrm{wt}(f) \leq \sum_{i=0}^{n-d-1} \binom{n}{i}.$$

Hence, for every n-variable function f, we have

$$\sum_{i=0}^{\mathcal{AI}_n(f)-1} \binom{n}{i} \leq \mathrm{wt}(f) \leq \sum_{i=0}^{n-\mathcal{AI}_n(f)} \binom{n}{i}.$$

Proof: Let g be a function of degree at most d. Let the ANF of g equal

$$a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{1 \leq i < j \leq n} a_{i,j} x_i x_j + \cdots + \sum_{1 \leq i_1 \leq \cdots \leq i_d \leq n} a_{i_1,\ldots i_d} x_{i_1} \ldots x_{i_d}.$$

Note that g is an annihilator of f if and only if $f(x) = 1$ implies $g(x) = 0$. Hence, g belongs to $AN(f)$ if and only if the coefficients in its ANF satisfy the system of homogeneous linear equations which translates this fact. In this system, we have $\sum_{i=0}^{d} \binom{n}{i}$ number of variables (the a's for the monomials up to degree d) and $\mathrm{wt}(f)$ many equations. If the number of variables is greater than the number of equations, then we will get nontrivial solutions. Thus, the fact that f has no annihilator g of degree d implies that the number of equations is greater than or equal to the number of variables, that is, $\mathrm{wt}(f) \geq \sum_{i=0}^{d} \binom{n}{i}$. Similarly, when considering $1+f$, we get $\mathrm{wt}(1+f) \geq \sum_{i=0}^{d} \binom{n}{i}$. This gives, $\mathrm{wt}(f) \leq 2^n - \sum_{i=0}^{d} \binom{n}{i}$, i.e., $\mathrm{wt}(f) \leq \sum_{i=0}^{n-d-1} \binom{n}{i}$. The last double inequality is obtained by choosing $d = \mathcal{AI}_n(f) - 1$. □

Theorem 1 gives an alternative proof of $\mathcal{AI}_n(f) \leq \lceil \frac{n}{2} \rceil$ which was given in [24]. Indeed, when applied to $d = \lceil \frac{n}{2} \rceil$, it leads to a contradiction, since we have

$$\sum_{i=0}^{\lceil \frac{n}{2} \rceil} \binom{n}{i} > \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - 1} \binom{n}{i}.$$

Note that the converse of Theorem 1 is not always true. For example, the affine functions are balanced, but clearly they have linear annihilators.

Applying Theorem 1 with $d = \lceil \frac{n}{2} \rceil - 1$, we get the following.

Corollary 1: $\mathcal{AI}_n(f) = \lceil \frac{n}{2} \rceil$ implies

1 f is balanced when n is odd;

2 $\sum_{i=0}^{\frac{n}{2}-1} \binom{n}{i} \leq \mathrm{wt}(f) \leq \sum_{i=0}^{\frac{n}{2}} \binom{n}{i}$ when n is even.

Now we connect algebraic immunity with nonlinearity. We first need a simple lemma, which has its own interest.

Lemma 1: For any $f \in B_n$ and any $l \in A_n$, we have

$$\mathcal{AI}_n(f) - 1 \leq \mathcal{AI}_n(f+l) \leq \mathcal{AI}_n(f) + 1. \qquad (1)$$

More generally, for any $f \in B_n$ and for any $h \in B_n$ whose algebraic degree equals r, we have

$$\mathcal{AI}_n(f) - r \leq \mathcal{AI}_n(f+h) \leq \mathcal{AI}_n(f) + r.$$

Proof: For any g such that $f * g = 0$, we have $(f+h) * ((h+1) * g) = 0$. For any g such that $(1+f) * g = 0$, we have $(1+f+h) * ((h+1) * g) = 0$. This gives the inequalities on the right. Applying them to $f+l$ and $f+h$ instead of f gives then the inequalities on the left. □

Note that these relations are still valid (changing n into the global number of variables) if f and l (respectively, h) are defined on different (maybe intersecting) sets of variables. Note also that, if these sets of variables are disjoint, then, denoting by m the global number of variables, we have

$$\mathcal{AI}_n(f) \leq \mathcal{AI}_m(f+l) \leq \mathcal{AI}_n(f) + 1$$

and

$$\mathcal{AI}_n(f) \leq \mathcal{AI}_m(f+h) \leq \mathcal{AI}_n(f) + r$$

since it is then possible to obtain an annihilator of degree $\mathcal{AI}_m(f+l)$ (respectively, $\mathcal{AI}_m(f+h)$) of f or $f+1$ by restricting to $F_2^n \times \{0\}$ an annihilator of the same degree of $f+l$ (respectively, $f+h$).

Siegenthaler [46] proposed to add to a given function f a linear function on disjoint variables for increasing the resiliency order of f; clearly, this secondary construction does not permit achieving good algebraic immunity.

Theorem 2: If $nl(f) < \sum_{i=0}^{d} \binom{n}{i}$, then $\mathcal{AI}_n(f) \leq d + 1$. More generally, if the Hamming distance $nl_r(f)$ between f and the set of Boolean functions of algebraic degrees at most r (the so-called Reed–Muller code of order r, $\mathrm{RM}(r,n)$) satisfies $nl_r(f) < \sum_{i=0}^{d} \binom{n}{i}$, then $\mathcal{AI}_n(f) \leq d + r$. In other words

$$nl_r(f) \geq \sum_{i=0}^{\mathcal{AI}_n(f)-r-1} \binom{n}{i}.$$

Proof: Let h be a function of degree at most r such that $nl_r(f) = d(f,h) = \mathrm{wt}(f+h)$. If $nl_r(f) < \sum_{i=0}^{d} \binom{n}{i}$ then $\mathcal{AI}_n(f+h) \leq d$, according to Theorem 1. Lemma 1 shows then that $\mathcal{AI}_n(f) \leq d + r$. The last inequality is obtained by choosing $d = \mathcal{AI}_n(f) - r - 1$. □

During the review process of this paper, a bound on the (first-order) nonlinearity has been obtained in [36]:

$$nl(f) \geq 2 \sum_{i=0}^{\mathcal{AI}_n(f)-2} \binom{n-1}{i}.$$

This bound improves upon the corresponding bound of Theorem 2. It has been further generalized in [18] to a bound on the higher order nonlinearity, which improves in some cases upon the corresponding bound of Theorem 2.

Theorem 2 and the result of [36] give a new reason why one should not use functions f with low nonlinearity, since in that case $\mathcal{AI}_n(f)$ would be low. However, they do not assure that if f has high algebraic immunity (for instance, an optimum one $\mathcal{AI}_n(f) = \lceil \frac{n}{2} \rceil$) then its nonlinearity will be high. Indeed, the result of [36] implies then that f has nonlinearity at least $2\sum_{i=0}^{\lceil\frac{n}{2}\rceil-2}\binom{n-1}{i}$, that is,

$$2^{n-1}-\binom{n-1}{\frac{n-1}{2}}$$

if n is odd and

$$2^{n-1}-\binom{n-1}{\frac{n}{2}-1}-\binom{n-1}{\frac{n}{2}}$$

if n is even. According to Stirling's formula, these values are approximately equal to $2^{n-1}-\frac{2^n}{\sqrt{2n\pi}}$ for odd n and to $2^{n-1}-\frac{2^{n+1}}{\sqrt{2n\pi}}$ for even n. They are very far from the maximum possible nonlinearity $2^{n-1}-2^{n/2-1}$.

A. Count of Annihilators

In the proof of Theorem 1, we get $\mathrm{wt}(f)$ many homogeneous linear equations whose variables are the coefficients in the ANF of g. Let us denote the coefficient matrix of this system of equations by $M_{f,d}$. Then $M_{f,d}$ has $\mathrm{wt}(f)$ many rows and $\sum_{i=0}^{d}\binom{n}{i}$ many columns. The rank $r_{f,d}$ of $M_{f,d}$ satisfies

$$r_{f,d}\le\min\left\{\mathrm{wt}(f),\sum_{i=0}^{d}\binom{n}{i}\right\}.$$

1) If $r_{f,d}=\sum_{i=0}^{d}\binom{n}{i}$, then there is no nonzero annihilator of degree $\le d$.
2) If $r_{f,d}<\sum_{i=0}^{d}\binom{n}{i}$, then there are nonzero annihilators of degree $\le d$. There will be $\sum_{i=0}^{d}\binom{n}{i}-r_{f,d}$ many linearly independent annihilators having degree $\le d$.

It is clear [24] that a larger number of independent annihilators helps better in cryptanalysis. Thus, when considering a Boolean function, one should check the number of independent annihilators at the lowest possible degree.

Definition 3: Given $f\in B_n$, we denote by $\#LDA_n(f)$ the number of independent annihilators of f of degree $\mathcal{AI}_n(f)$.

Theorem 3:

1) Take $f\in B_n$, with $\mathcal{AI}_n(f)=d$. Then $\#LDA_n(f)\le\binom{n}{d}$.
2) Take balanced $f\in B_n$ with $\mathcal{AI}_n(f)=\frac{n}{2}$, n even. Then $\#LDA_n(f)\ge\frac{1}{2}\cdot\binom{n}{\frac{n}{2}}$.
3) Take $f\in B_n$ such that $\mathcal{AI}_n(f)=\frac{n+1}{2}$, n odd. Then $\#LDA_n(f)=\binom{n}{\frac{n+1}{2}}$.

Proof: The proof of item 1) is as follows: if two annihilators of degree d have the same degree d part in their algebraic normal forms, then they must be equal since their sum being then an annihilator of degree strictly smaller than d, it must be null. We deduce that $\#LDA_n(f)$ is upper-bounded by the dimension of the quotient $\mathrm{RM}(d,n)/\mathrm{RM}(d-1,n)$, that is, $\binom{n}{d}$.

Now we prove item 2). Here, $\mathrm{wt}(f)=2^{n-1}$. The function f has an annihilator of degree $\frac{n}{2}$. The corresponding coefficient matrix $M_{f,\frac{n}{2}}$ has 2^{n-1} many rows and

$$\sum_{i=0}^{\frac{n}{2}}\binom{n}{i}=2^{n-1}+\frac{1}{2}\cdot\binom{n}{\frac{n}{2}}$$

many columns. Thus, rank of $M_{f,\frac{n}{2}}$ is at most 2^{n-1}. The number of independent solutions is lower bounded by

$$\left(2^{n-1}+\frac{1}{2}\cdot\binom{n}{\frac{n}{2}}\right)-2^{n-1}=\frac{1}{2}\cdot\binom{n}{\frac{n}{2}}.$$

Now we prove item 3). Here $\mathrm{wt}(f)=2^{n-1}$, according to Corollary 1. By hypothesis, there is no nonzero annihilator up to degree $\frac{n-1}{2}$. The coefficient matrix $M_{f,\frac{n-1}{2}}$ is a $2^{n-1}\times 2^{n-1}$ square matrix, since

$$\sum_{i=0}^{\frac{n-1}{2}}\binom{n}{i}=2^{n-1}.$$

As it has no nontrivial solution, its rank r equals 2^{n-1}. The function f has an annihilator of degree $\frac{n+1}{2}$. In this case, the corresponding coefficient matrix $M_{f,\frac{n+1}{2}}$ has 2^{n-1} many rows and $2^{n-1}+\binom{n}{\frac{n+1}{2}}$ many columns. Thus, the rank of $M_{f,\frac{n+1}{2}}$ equals that of $M_{f,\frac{n-1}{2}}$, i.e., equals 2^{n-1}. The number of independent solutions equals

$$\left(2^{n-1}+\binom{n}{\frac{n+1}{2}}\right)-2^{n-1}=\binom{n}{\frac{n+1}{2}}.\qquad\square$$

In Section IV, we study certain constructions of cryptographically significant Boolean functions in terms of algebraic immunity.

IV. Studying Functions for Their Algebraic Immunity

A statistical analysis has shown in [38] that any randomly chosen balanced function on large number of variables has no bad algebraic immunity with very high probability. This result has the same flavor as the fact that most of the Boolean functions have high algebraic degrees and high nonlinearities in general (see [39]). That is, if one chooses a Boolean function randomly, the probability that these three characteristics will not be bad is high. Heuristic arguments exposed in [20] suggest even that almost all Boolean functions have in fact algebraic immunity at least $\lfloor\frac{n}{2}\rfloor$. This has been later confirmed in [51]. However, when considering a specific construction technique, the number of functions constructed by that method is much lower than the total space of Boolean functions and generally such statistical analysis does not work.

A. Experimental Results on Rotation Symmetric Boolean Functions

If we intend to construct (n,m,d,x) functions with best possible parameters along with the best possible algebraic immunity, we can first consider a subset of Boolean functions, which is sufficiently particular so that the study will be simplified (mathematically and/or algorithmically) and sufficiently nonpeculiar so that it will be possible to find such functions (as it can be with random ones). The rotation symmetric Boolean functions (RSBFs) received a lot of attention recently for this reason [32], [47], [48], [33], [37]. These functions are invariant under circular translation of indices in the input variables. We present experimental results related to the algebraic immunity of the RSBFs which are available in [47], [48], [33], [37].

Experiment 1: Here we test the algebraic immunity for $(7, 2, 4, 56)$ RSBFs. It is given in [47] that there are 36 such functions with $f(0) = 0$. Out of them, 24 functions contain linear terms. For these functions, $\mathcal{AI}_n(f)$ equals 3, which is one less than the highest possible value $\lceil \frac{n}{2} \rceil = 4$. Out of them 12 functions have $\#LDA_n(f) = 3$ and the remaining 12 have $\#LDA_n(f) = 4$. The 12 functions having no linear term have algebraic immunity $\mathcal{AI}_n(f) = 4$, which is the highest possible value. According to Theorem 3 (item 3) (we have also checked this by experiment), for these functions $\#LDA_n(f) = \binom{7}{\lceil \frac{7}{2} \rceil} = 35$.

Experiment 2: Here we examine the $(8, 1, 6, 116)$ RSBFs with $f(0) = 0$ the number of which is 10272 as recalled in [48]. Out of them, 6976 attain highest algebraic immunity, i.e., 4 and we find that for these functions $\#LDA_n(f) = 35$. Theorem 3 (item 2) asserts that the value should be greater than or equal to $\frac{\binom{8}{\frac{8}{2}}}{2} = 35$. This gives an example, where the bound is tight. For the remaining $10272 - 6976 = 3296$ functions, the algebraic immunity is 3. Out of them, 1536 many functions f have only one annihilator of degree 3 (but no degree 3 annihilator for $1 + f$), 1504 many functions f have no annihilator of degree 3 (but one degree 3 annihilator for $1 + f$), and 256 many functions f have one annihilator of degree 3 and also one degree 3 annihilator for $1 + f$. According to Theorem 3 (item 1), $\#LDA_n(f) \leq \binom{8}{3} = 56$. So for these functions, the bound is not sharp.

Experiment 3: In the preceding two experiments, we examined the functions which are balanced. Now we consider the $[9, 3, 5, 240]$ RSBFs which are not balanced. We consider the 8406 functions with $f(0) = 0$, see [33], [37]. According to Corollary 1 (item 1), the algebraic immunity of these functions will be strictly less than 5. Here, after experiment, we get the algebraic immunity of all 8406 functions as 4. According to Theorem 3 (item 1), $\#LDA_9(f) \leq \binom{9}{4} = 126$. In the table at the bottom of the page, we present the number of functions satisfying a particular $\#LDA_9(f)$ and $\#LDA_9(1 + f)$.

Studying the resilient functions on seven and eight variables and unbalanced correlation immune functions on nine variables for this rotation-symmetric class of Boolean functions, it is evident that there exist functions which are good in terms of algebraic immunity.

B. Analysis of Some Construction Methods

Very few primary constructions of Boolean functions achieving at high levels the cryptographic criteria recalled in the introduction are known (see [19]). A general principle of construction exists: concatenating low-degree functions as in the Maiorana–McFarland construction. But this principle has some limits with respect to the usual criteria (see [14]) and it has drawbacks with respect to the algebraic immunity as we show now.

1) The Maiorana–McFarland Construction: The original Maiorana–McFarland class of bent functions is as follows (see e.g., [13]). Consider n-variable Boolean functions of the form $f(x, y) = x \cdot \pi(y) + g(y)$, where $x, y \in F_2^{\frac{n}{2}}$, π is a permutation on $F_2^{\frac{n}{2}}$ and g is any Boolean function on $\frac{n}{2}$ variables. Function f can be seen as concatenation of $2^{\frac{n}{2}}$ distinct (up to complementation) affine functions on $\frac{n}{2}$ variables.

A similar type of concatenation technique has also been used for construction of resilient functions [11] (see also [44], [42]). Concatenating k-variable affine functions (with repetition allowed) nondegenerate on at least $m + 1$ variables generates an m-resilient function f on n-variables. For such a function f, it is easy to find an annihilator of degree $n - k + 1$ as described in [38]. In fact, it is shown in [20] that, unless a heavy condition is satisfied (which is very improbable unless k is almost equal to n), it is easy to find an annihilator of degree $n - k$. It has been commented in [38, Example 1 and the following paragraph] that k is generally greater than $\frac{n}{2}$ (this seems true for the Maiorana–McFarland type of functions presented in [41], [14]; but this has not been checked for some large classes of Maiorana–McFarland type of functions described in [42], [17]) and hence it is possible to get a nonzero annihilator g of degree less than $\frac{n}{2}$. However, it should be noted that in construction of resilient functions, there are techniques that use concatenation of k-variable affine functions where $k < \frac{n}{2}$. In such a case, the annihilators described above will have degree greater than $\frac{n}{2}$ and will not be of practical use as there are other annihilators of degree $\leq \frac{n}{2}$ which are not of the form given in [38, Theorem 2].

2) Secondary Constructions:

- We first study a construction of functions proposed by Siegenthaler [46]. Given $f \in B_n$, we denote by $LDGA_n(f)$ the set of non null $f_1 \in B_n$ with lowest possible degree such that $f * f_1 = 0$ or $(1 + f) * f_1 = 0$.

Proposition 1: Let f, g be two Boolean functions on the variables $x_1, x_2, \ldots, x_n$ with $\mathcal{AI}_n(f) = d_1$ and $\mathcal{AI}_n(g) = d_2$. Let $h = (1 + x_{n+1})f + x_{n+1}g \in B_{n+1}$. Then

1) If $d_1 \neq d_2$ then $\mathcal{AI}_{n+1}(h) = \min\{d_1, d_2\} + 1$.
2) If $d_1 = d_2 = d$, then $d \leq \mathcal{AI}_{n+1}(h) \leq d + 1$, and $\mathcal{AI}_{n+1}(h) = d$ if and only if there exists $f_1, g_1 \in B_n$ of algebraic degree d such that $\{f * f_1 = 0, g * g_1 = 0\}$ or $\{(1 + f) * f_1 = 0, (1 + g) * g_1 = 0\}$ and $\deg(f_1 + g_1) \leq d - 1$.

$\#LDA_9(f)$	16	17	18	19	20	21
$\#LDA_9(1+f)$	0	1	2	3	4	5
$\#f$	5658	1758	774	180	12	24

Proof: Let $f_1 \in LDGA_n(f)$ and $g_1 \in LDGA_n(g)$. If $f * f_1 = 0$, then $(1 + x_{n+1}) * f_1 * h = 0$. If $(1 + f) * f_1 = 0$, then $(1 + x_{n+1}) * f_1 * (1 + h) = 0$. Also, if $g * g_1 = 0$, then $x_{n+1} * g_1 * h = 0$ and if $(1+g) * g_1 = 0$, then $x_{n+1} * g_1 * (1+h) = 0$. Thus,

$$\mathcal{AI}_{n+1}(h) \le \min\{\mathcal{AI}_n(f), \mathcal{AI}_n(g)\} + 1. \tag{2}$$

Let $p = (1 + x_{n+1})p_1 + x_{n+1}p_2 \in LDGA_{n+1}(h)$. Let us first consider the case with $h * p = 0$ which implies

$$(1 + x_{n+1})f * p_1 + x_{n+1}g * p_2 = 0.$$

So $f * p_1 = 0$ and $g * p_2 = 0$. Similarly, for the case with $(1 + h) * p = 0$, i.e.,

$$(1 + x_{n+1}) * (1 + f) * p_1 + x_{n+1}(1 + g) * p_2 = 0$$

we have $(1 + f) * p_1 = 0$ and $(1 + g) * p_2 = 0$. Now there can be three cases in both scenarios.

a) p_1 is zero and p_2 is nonzero. So $\deg(p_2) \ge d_2$ which gives $\deg(p) \ge d_2 + 1$.

b) p_2 is zero and p_1 is nonzero. So $\deg(p_1) \ge d_1$ which gives $\deg(p) \ge d_1 + 1$.

c) Both p_1, p_2 are nonzero. So $\deg(p_1) \ge d_1$ and $\deg(p_2) \ge d_2$, which gives $\deg(p) \ge \max\{d_1, d_2\} + 1$, when $d_1 \ne d_2$.

So for $d_1 \ne d_2$ we get

$$\mathcal{AI}_{n+1}(h) \ge \min\{\mathcal{AI}_n(f), \mathcal{AI}_n(g)\} + 1. \tag{3}$$

Equations (2) and (3) give the proof of item 1).

Now we prove item 2). Consider

$$p = (1+x_{n+1})f_1 + x_{n+1}g_1 = f_1 + x_{n+1}(f_1 + g_1) \in LDGA_{n+1}(h).$$

Clearly, p has degree at least d, since f_1 has degree at least d. So, $d \le \mathcal{AI}_{n+1}(h) \le d + 1$.

If $\mathcal{AI}_{n+1}(h) = d$, then the highest degree terms of f_1 and g_1 must be same which gives $deg(f_1 + g_1) \le d - 1$. Note that we have $\{f * f_1 = 0, g * g_1 = 0\}$ or $\{(1 + f) * f_1 = 0, (1 + g) * g_1 = 0\}$. Conversely, if there exists $f_1, g_1 \in B_n$ of algebraic degree d such that $\{f * f_1 = 0, g * g_1 = 0\}$ or $\{(1 + f) * f_1 = 0, (1 + g) * g_1 = 0\}$ and $\deg(f_1 + g_1) \le d - 1$, then clearly $\mathcal{AI}_{n+1}(h) = d$. □

We cannot say that the construction of Proposition 1, first introduced by Siegenthaler [46] for obtaining resilient functions, is good or is bad in terms of algebraic immunity, since

— a good construction is supposed to gain 1 (respectively, k) for the algebraic immunity when we add 2 (respectively, $2k$) variables, here we add only one;

— the construction is very general since every function can be obtained from it.

The next corollary is a direct consequence of Proposition 1 and of the upper bound $\lceil \frac{n}{2} \rceil$ on the algebraic immunity of n-variable functions.

Corollary 2: Let

$$h = (1 + x_{n+1})f + x_{n+1}g \in B_{n+1}$$

where n is even and $\mathcal{AI}_{n+1}(h) = \frac{n}{2} + 1$ (i.e., has maximum possible value). Then $\mathcal{AI}_n(f) = \mathcal{AI}_n(g) = \frac{n}{2}$ (i.e., is maximum) and there do not exist $f_1, g_1 \in B_n$ of degree $\frac{n}{2}$ such that $\{f * f_1 = 0 \text{ and } g * g_1 = 0\}$ or $\{(1+f) * f_1 = 0 \text{ and } (1+g) * g_1 = 0\}$ and such that all $\frac{n}{2}$ degree monomials of f_1 and g_1 are the same. We now observe that two functions on an odd number n of variables and with optimum algebraic immunity always have some relationships.

Corollary 3: Let $f, g \in B_n$ where n is odd and $\mathcal{AI}_n(f) = \mathcal{AI}_n(g) = \frac{n+1}{2}$ (the maximum possible value). Then there must exist $f_1, g_1 \in B_n$ of degree $\frac{n+1}{2}$ such that $\{f * f_1 = 0$ and $g * g_1 = 0\}$ or $\{(1 + f) * f_1 = 0$ and $(1 + g) * g_1 = 0\}$ and such that all $\frac{n+1}{2}$ degree monomials of f_1 and g_1 are same.

Proof: Let

$$h = (1 + x_{n+1})f + x_{n+1}g \in B_{n+1}.$$

According to Proposition 1, $\mathcal{AI}_{n+1}(h)$ equals $\frac{n+1}{2}$ since it cannot be greater than $\frac{n+1}{2}$. □

• In [49], Tarannikov has proposed an important construction of resilient functions. A similar kind of construction has been derived in [40] (and has been later generalized in [15]). It has been shown in [27] that if we denote by H_0 the function from which we start in this construction and by H_i the function obtained after i steps (this function has $3i$ more variables than H_0), then $\mathcal{AI}_n(H_0) \le \mathcal{AI}_{n+3i}(H_i) \le \mathcal{AI}_n(H_0) + i + 2$. Later, it has been proved in [7] that the n-variable functions constructed by Tarannikov's method [49], [40] attain $\mathbf{\Omega}(\sqrt{n})$ algebraic immunity.

• We also like to present some observations on $(9, 1, 7, 240)$ functions constructed in [42, Theorem 10(b)]. The operation on strings $x\$y = (x \otimes y^c) + (x^c \otimes y)$ is defined in [42], where x^c, y^c are bitwise complements of x, y, respectively, and where $\otimes$ is the Kronecker product, whose definition is: $(x_i)_{i \in I} \otimes (y_j)_{j \in J} = (x_i y_j)_{(i,j) \in I \times J}$. Obviously, when applied to strings corresponding to the truth tables of Boolean functions, this operation \$ corresponds to the so-called *direct sum*, that is, the addition of Boolean functions with disjoint sets of variables (if the definitions of the functions use same symbols to designate some variables, then these symbols must be duplicated so that the functions become defined on different variables). Now we present the construction of a $(2p + 1, 1, 2p - 1, 2^{2p} - 2^p)$ function as given in [42] for $p \ge 4$.

Construction 1: [42, Theorem 10(b)] Let λ_1, λ_2, λ_3, λ_4 be the 3-variable linear functions nondegenerate on two variables (i.e., the functions $x_1 + x_2$, $x_2 + x_3$, $x_1 + x_3$, $x_1 + x_2 + x_3$). Let g_i be the 4-variable function $x_i + x_4$, for $i = 1, 2, 3$. Let h_1, h_2 be bent functions on $(2p - 4)$ variables, let h_3, h_4, h_5 be bent functions of $(2p - 6)$ variables and h_6, h_7 be two strings of lengths $2^{2p-6} + 1$ and $2^{2p-6} - 1$ which are prepared by properly adding and removing 1 bit from the truth table of $(2p - 6)$-variable bent functions, respectively. Let f be a concatenation of the following sequence of functions. $h_1\$\lambda_1$, $h_2\$\lambda_2$, $h_3\$g_1$, $h_4\$g_2$, $h_5\$g_3$, $h_6\$\lambda_3$, $h_7\$\lambda_4$. This is a $(2p + 1, 1, 2p - 1, 2^{2p} - 2^p)$ function.

Example 1: For $p = 4$, we choose: $h_1 = 0000010100110110$, $h_2 = 0000010100110110$, $h_3 = 0001$,

$h_4 = 0001$, $h_5 = 0001$, $h_6 = 00010$, $h_7 = 001$. In this case, we find a $(9,1,7,240)$ function f_1 with $\mathcal{AI}_9(f_1) = 3$. If one replaces the function $h_2 = 0000010100110110$ by $h_2 = 0000010100111001$, then we get a $(9,1,7,240)$ function f_2 with $\mathcal{AI}_9(f_2) = 4$.

We observed that changing the order of affine functions can change the algebraic immunity without any change in order of resiliency, nonlinearity, and algebraic degree. The change in the last four bits in h_2 implies that the concatenation of λ_2, $1+\lambda_2$, $1+\lambda_2$, λ_2 will be replaced by $1+\lambda_2$, λ_2, λ_2, $1+\lambda_2$. We observed that this increases the algebraic immunity from 3 to 4.

C. *Functions With Low-Degree Subfunctions*

In this subsection, we discuss why a Boolean function with low-degree subfunction is not good in terms of algebraic immunity. This extension of a result presented in [38], and its complements, are simple, but they have some importance for the design of pseudorandom generators.

Proposition 2: Let $f \in B_n$. Let $g \in B_{n-r}$ be a subfunction of $f(x_1,\ldots,x_n)$ after fixing r many distinct inputs $x_{i_1},\ldots,x_{i_r} \in \{x_1,\ldots,x_n\}$. If the algebraic degree of g is d, then $\mathcal{AI}_n(f) \le d+r$.

Proof: Let $x_{i_1},\ldots,x_{i_r}$ be fixed at the values $a_{i_1},\ldots,a_{i_r} \in F_2$. Thus, g is a function on the variables $\{x_1,\ldots,x_n\} \setminus \{x_{i_1},\ldots,x_{i_r}\}$. Obviously, $(1+a_{i_1}+x_{i_1})\ldots(1+a_{i_r}+x_{i_r})(1+g)$ is an annihilator of f. The algebraic degree of $(1+a_{i_1}+x_{i_1})\ldots(1+a_{i_r}+x_{i_r})(1+g)$ is $d+r$. □

The Maiorana–McFarland construction can be seen as concatenation of 2^r affine functions on $n-r$ variables to construct an n-variable function. Clearly, we have affine subfunctions of the constructed function in this case, and hence $\deg(g) = 1$ following the notation of Proposition 2. Thus, as already recalled at Section IV-B1, there will be annihilators of degree $1+r$. Note that if r is small, then one can get annihilators at low degree [38, Theorem 2, Example 1]. This works for any function, which needs not be of Maiorana–McFarland type only. For instance, let us consider a 20-variable function, with a subfunction of degree 2 on 17-variables, i.e., we fix three inputs. In that case, the 20-variable function will have an annihilator of degree $2+3=5$.

Proposition 3: The $(2p+1)$-variable function presented in Construction 1 has a subfunction of degree at most $p-1$ when $x_{2p+1} = 0$.

Proof: Consider the subfunction when $x_{2p+1} = 0$. The subfunction (call it g) in concatenation form is $h_1\$\lambda_1$, $h_2\$\lambda_2$. Since h_1, h_2 are bent functions on $2p-4$ variables, they can have algebraic degree at most $p-2$. Further, λ_1, λ_2 are 3-variable linear functions. The algebraic normal form of g is $(1+x_{2p})(h_1+\lambda_1) + x_{2p}(h_2+\lambda_2)$. So the degree of g is smaller or equal to $1+(p-2) = p-1$. □

Theorem 4: For a function $f \in B_n$ (n odd) generated out of Construction 1, $\mathcal{AI}_n(f) \le \lfloor \frac{n}{2} \rfloor$.

Proof: Here $n = 2p+1$. We take $g \in B_{n-1}$, i.e., $r = 1$ according to Proposition 2. Further from Proposition 3

$$\deg(g) \le p-1 = \frac{n-1}{2} - 1.$$

Thus, $\mathcal{AI}_n(f) \le \frac{n-1}{2} - 1 + 1 = \lfloor \frac{n}{2} \rfloor$. □

Now we answer why the algebraic immunity of these two functions in Example 1 are different. The reason is that, in the first case, the functions h_1, h_2 are same with the ANF $x_1x_3 + x_2x_4$. Thus, the subfunction g (i.e., $h_1\$\lambda_1$, $h_2\$\lambda_2$) is a degree–2 function. So the maximum algebraic immunity, according to Proposition 2, can be $2+1=3$. In the second case, h_1 is different from h_2 and the algebraic degree of g (i.e., $h_1\$\lambda_1$, $h_2\$\lambda_2$) becomes 3 and it achieves the value $3+1=4$. Thus, Proposition 2 helps in answering this question. It is important to note that this technique can be employed to study the upper bound of algebraic immunity for various constructions by analyzing their subfunctions and in particular, directly for the constructions proposed in [42], [14].

It should be noted that the converse of Proposition 2 is not always true. That is, a function having low-degree annihilator does not need to have some low-degree subfunction by fixing a few variables. As example, one may refer to the 5-variable function

$$f = x_1 + x_2 + x_2x_4 + x_3x_4 + (x_2 + x_3 + x_1x_4 + x_2x_4 + x_3x_4)x_5.$$

This function has algebraic immunity 2 and the only annihilator of degree 2 is

$$1 + x_1 + x_2 + x_1x_4 + x_3x_4 + (x_2 + x_3 + x_4)x_5.$$

If one verifies all possible subfunctions of f after fixing 1 and 2 variables, it is not possible to get subfunctions of degree 1 and 0, respectively.

Note that the observation we made for Maiorana–McFarland's functions does not seem to apply to those Boolean functions that can be seen as concatenations of indicators of flats [17].

V. Construction to Get Optimal Algebraic Immunity

We have recalled in Section IV-B that very few primary constructions of Boolean functions achieving at high levels the usual cryptographic criteria are known, and we have seen that these constructions do not seem to be able to achieve good algebraic immunity. In this section, we present a construction to design a Boolean function of $2k$ variables with algebraic immunity k. The construction is iterative in nature. At each step, two variables are added and the algebraic immunity is increased by 1. The constructed function is not balanced, but the bias with respect to balancedness tends to zero when k tends to infinity. The constructed function has not a high nonlinearity either. The bias with respect to optimum nonlinearity is slightly better than the minimum observed in Section III after Theorem 2. This primary construction can be (must be) combined with secondary constructions to lead to functions satisfying all of the necessary cryptographic criteria. Since we will be able to give a very efficient way of computing its output (with a linear complexity in the number of variables), its introduction in a secondary construction is efficient even if its number of variables is large. Last but not least argument, it is the first known provably efficient way of obtaining functions with optimal algebraic immunity (the next one appeared in [29]). We show that the function has very high algebraic degree.

Construction 2: We denote by $\phi_{2k} \in B_{2k}$ the function defined by the recursion

$$\phi_{2k+2} = \phi_{2k}||\phi_{2k}||\phi_{2k}||\phi^1_{2k} \tag{4}$$

where $||$ denotes the concatenation, (in terms of algebraic normal form, we have then $\phi_{2k+2} = \phi_{2k} + x_{2k+1}x_{2k+2}(\phi_{2k} + \phi^1_{2k})$), and where ϕ^1_{2k} is defined itself by a doubly indexed recursion

$$\phi^i_{2j} = \phi^{i-1}_{2j-2}||\phi^i_{2j-2}||\phi^i_{2j-2}||\phi^{i+1}_{2j-2} \tag{5}$$

i.e., in terms of algebraic normal form

$$\begin{aligned}\phi^i_{2j} = \phi^{i-1}_{2j-2} &+ (x_{2j-1} + x_{2j})\left(\phi^{i-1}_{2j-2} + \phi^i_{2j-2}\right)\\ &+ x_{2j-1}x_{2j}\left(\phi^{i-1}_{2j-2} + \phi^{i+1}_{2j-2}\right), \quad \text{for } j > 0, i > 0\end{aligned}$$

with base step $\phi^0_j = \phi_j$ for $j > 0$, $\phi^i_0 = i \mod 2$ for $i \geq 0$.

To understand the recursion in the Construction 2, we present an example up to some depth.

- $\phi^1_{2k} = \phi_{2k-2}||\phi^1_{2k-2}||\phi^1_{2k-2}||\phi^2_{2k-2}$.
- $\phi^2_{2k-2} = \phi^1_{2k-4}||\phi^2_{2k-4}||\phi^2_{2k-4}||\phi^3_{2k-4}$.
- $\phi^3_{2k-4} = \phi^2_{2k-6}||\phi^3_{2k-6}||\phi^3_{2k-6}||\phi^4_{2k-6}$.

This goes on until we reach the null level for at least one of the two indices.

Below we present the construction idea as truth table concatenation.

Step 1: $\phi_2 = 0001$
Step 2: $\phi_4 = \phi_2\phi_2\phi_2 0110$
Step 3: $\phi_6 = \phi_4\phi_4\phi_4\phi_2 011001101001$
Step 4:

$$\begin{aligned}\phi_8 = \phi_6\phi_6\phi_6\phi_4\phi_2 &011001101001\\ &\phi_2 0110011010010110100110010110.\end{aligned}$$

To prove that ϕ_{2k} has algebraic immunity k, we need intermediate results. In the proofs, we will use the fact that, for any $f \in B_n$ and any subset V of $\{0,1\}^n$, the restriction to V of an annihilator of f is an annihilator of the restriction of f to V. For technical reasons, during our proofs, we will encounter certain situations when the degree of a function is negative. As such functions cannot exist, we will replace those functions by function 0.

Lemma 2: Assume that the function $\phi_{2i} \in B_{2i}$ has been generated by Construction 2 for $0 \leq i \leq k$ and that $\mathcal{AI}_{2i}(\phi_{2i}) = i$ for $0 \leq i \leq k$. If, for some $0 \leq i \leq k$ and $j \geq 0$, there exist $g \in AN(\phi^j_{2i})$ and $h \in AN(\phi^{j+1}_{2i})$ such that $\deg(g+h) \leq i-2-j$ then $g = h$.

Proof: We prove Lemma 2 by induction on i.

For the base step $i = 0$, $\deg(g+h) \leq 0-2-j \leq -2$ implies that such a function cannot exist, i.e., $g+h$ is identically 0, which gives $g = h$.

Now we prove the inductive step. Assume that, for $i < \ell$, the induction assumption holds (for every $j \geq 0$). We will show it for $i = \ell$ (and for every $j \geq 0$). Suppose that there exist $g \in AN(\phi^j_{2\ell})$ and $h \in AN(\phi^{j+1}_{2\ell})$ with $\deg(g+h) \leq \ell-2-j$. By construction, if $j > 0$ then we have

$$\begin{aligned}\phi^j_{2\ell} &= \phi^{j-1}_{2(\ell-1)}||\phi^j_{2(\ell-1)}||\phi^j_{2(\ell-1)}||\phi^{j+1}_{2(\ell-1)}\\ \phi^{j+1}_{2\ell} &= \phi^j_{2(\ell-1)}||\phi^{j+1}_{2(\ell-1)}||\phi^{j+1}_{2(\ell-1)}||\phi^{j+2}_{2(\ell-1)}\end{aligned}$$

and if $j = 0$ then

$$\phi^0_{2\ell} = \phi^0_{2(\ell-1)}||\phi^0_{2(\ell-1)}||\phi^0_{2(\ell-1)}||\phi^1_{2(\ell-1)}.$$

Let us denote

$$\begin{aligned}g &= v_1||v_2||v_3||v_4,\\ h &= v_5||v_6||v_7||v_8.\end{aligned}$$

Since $\deg(g+h) \leq \ell-2-j$, from the ANF of $g+h = (v_1+v_5) + x_{2\ell-1}(v_1+v_5+v_2+v_6) + x_{2\ell}(v_1+v_5+v_3+v_7) + x_{2\ell-1}x_{2\ell}(v_1+\cdots+v_8)$ we deduce the following.

- $\deg(v_1+v_5) \leq \ell-2-j = (\ell-1)-2-(j-1)$. If $j > 0$ then $v_1 \in AN(\phi^{j-1}_{2(\ell-1)})$, $v_5 \in AN(\phi^j_{2(\ell-1)})$ implies that $v_1 = v_5$, according to the induction assumption. If $j = 0$, then we have $v_1, v_5 \in AN(\phi_{2(\ell-1)})$, and therefore $(v_1 + v_5) \in AN(\phi_{2(\ell-1)})$, with $\deg(v_1+v_5) \leq \ell-2$. Suppose that $v_1 + v_5 \neq 0$, then we would have $\deg(v_1+v_5) \geq \ell-1$, since $\mathcal{AI}_{2(\ell-1)}(\phi_{2(\ell-1)}) = \ell-1$, by hypothesis; a contradiction. Hence $v_1 + v_5 = 0$ i.e., $v_1 = v_5$.
- $\deg(v_2+v_6) \leq (\ell-1)-2-j$ and $v_2 \in AN(\phi^j_{2(\ell-1)})$, $v_6 \in AN(\phi^{j+1}_{2(\ell-1)})$, imply that $v_2 = v_6$, according to the induction assumption.
- $\deg(v_3+v_7) \leq (\ell-1)-2-j$ and $v_3 \in AN(\phi^j_{2(\ell-1)})$, $v_7 \in AN(\phi^{j+1}_{2(\ell-1)})$, imply that $v_3 = v_7$, according to the induction assumption.
- $\deg(v_4+v_8) \leq (\ell-1)-2-(j+1)$ and $v_4 \in AN(\phi^{j+1}_{2(\ell-1)})$, $v_8 \in AN(\phi^{j+2}_{2(\ell-1)})$, imply that $v_4 = v_8$, according to the induction assumption.

Hence, we get $g = h$. □

Lemma 3: Assume that the function $\phi_{2i} \in B_{2i}$ has been generated by Construction 2 for $0 \leq i \leq k$ and that $\mathcal{AI}_{2i}(\phi_{2i}) = i$ for $0 \leq i \leq k$. If, for some $0 \leq i \leq k$ and $j \geq 0$, there exists $g \in AN(\phi^j_{2i}) \cap AN(\phi^{j+1}_{2i})$ such that $\deg(g) \leq i+j$, then $g = 0$.

Proof: We prove Lemma 3 by induction on $i-j$.

For the base step (i.e., $i-j \leq 0$), we have from construction $\phi^{j+1}_{2i} = \phi^j_{2i}+1$ (this can easily be checked by induction). Hence, $g \in AN(\phi^j_{2i}) \cap AN(\phi^j_{2i}+1)$, and $g = 0$.

Now we prove the inductive step. Assume that the induction assumption holds for $i-j \leq \ell$, $\ell \geq 0$, and let us prove it for $i-j = \ell+1$. So let $g \in AN(\phi^j_{2i}) \cap AN(\phi^{j+1}_{2i})$ where $i-j = \ell+1$.

If $j > 0$, we have

$$\begin{aligned}\phi^j_{2i} &= \phi^{j-1}_{2(i-1)}||\phi^j_{2(i-1)}||\phi^j_{2(i-1)}||\phi^{j+1}_{2(i-1)},\\ \phi^{j+1}_{2i} &= \phi^j_{2(i-1)}||\phi^{j+1}_{2(i-1)}||\phi^{j+1}_{2(i-1)}||\phi^{j+2}_{2(i-1)}.\end{aligned}$$

Let us denote

$$g = v_1 \| v_2 \| v_3 \| v_4, \text{ we have}$$

$$\begin{aligned} v_1 &\in AN\left(\phi_{2(i-1)}^{j-1}\right) \cap AN\left(\phi_{2(i-1)}^{j}\right) \\ v_2, v_3 &\in AN\left(\phi_{2(i-1)}^{j}\right) \cap AN\left(\phi_{2(i-1)}^{j+1}\right) \end{aligned}$$

and

$$v_4 \in AN\left(\phi_{2(i-1)}^{j+1}\right) \cap AN\left(\phi_{2(i-1)}^{j+2}\right).$$

1) Since $\deg(g) \le i+j$, we have $\deg(v_4) \le i+j = (i-1)+(j+1)$. Since $(i-1)-(j+1) = i-j-2 < \ell$, we have $v_4 = 0$, according to the induction assumption. So the ANF of g is $v_1 + x_{2i-1}(v_1+v_2) + x_{2i}(v_1+v_3) + x_{2i-1}x_{2i}(v_1+v_2+v_3)$. Then $\deg(v_1+v_2)$, $\deg(v_1+v_3)$, $\deg(v_1+v_2+v_3) \le i+j-1$, which implies $\deg(v_1)$, $\deg(v_2)$, $\deg(v_3) \le i+j-1$.
2) We have then $\deg(v_2) \le i+j-1 = (i-1)+j$ and $\deg(v_3) \le i+j-1 = (i-1)+j$. Since $(i-1)-j = i-j-1 \le \ell$, we have $v_2 = v_3 = 0$, according to the induction assumption.
3) Since $v_2 = v_3 = v_4 = 0$, the ANF of g is $(1+x_{2i-1}+x_{2i}+x_{2i-1}x_{2i})v_1$. So, $\deg(v_1) \le i+j-2 = (i-1)+(j-1)$. Here $(i-1)-(j-1) = \ell+1$. So, we can not use the induction assumption directly. Now we break v_1 again into four parts as

$$\begin{aligned} \phi_{2(i-1)}^{j-1} &= \phi_{2(i-2)}^{j-2} \| \phi_{2(i-2)}^{j-1} \| \phi_{2(i-2)}^{j-1} \| \phi_{2(i-2)}^{j} \\ \phi_{2(i-1)}^{j} &= \phi_{2(i-2)}^{j-1} \| \phi_{2(i-2)}^{j} \| \phi_{2(i-2)}^{j} \| \phi_{2(i-2)}^{j+1} \\ v_1 &= v_{1,1} \| v_{1,2} \| v_{1,3} \| v_{1,4}. \end{aligned}$$

Using similar arguments as in items 1) and 2), we have $v_{1,2} = v_{1,3} = v_{1,4} = 0$. So, $\deg(v_{1,1}) \le i+j-4$. Doing the similar process j times, we will get some function $v \in AN(\phi_{2(i-j)}) \cap AN(\phi_{2(i-j)}^1)$. At every step of this subinduction, the degree decreases by 2, and we have then $\deg(v) \le i+j-2j = i-j$. Breaking v for the last time into four parts and using that $v \in AN(\phi_{2(i-j)}) \cap AN(\phi_{2(i-j)}^1)$, we have

$$\begin{aligned} \phi_{2(i-j)} &= \phi_{2(i-j-1)} \| \phi_{2(i-j-1)} \| \phi_{2(i-j-1)} \| \phi_{2(i-j-1)}^1 \\ \phi_{2(i-j)}^1 &= \phi_{2(i-j-1)} \| \phi_{2(i-j-1)}^1 \| \phi_{2(i-j-1)}^1 \| \phi_{2(i-j-1)}^2 \\ v &= v' \| v'' \| v''' \| v''''. \end{aligned}$$

Using similar arguments as in items 1) and 2), we have $v'' = v''' = v'''' = 0$. So, $\deg(v') \le i-j-2$. And $v' \in AN(\phi_{2(i-j-1)})$ implies that, if $v' \ne 0$, then $\deg(v) \ge i-j-1$, a contradiction. Hence, $v' = 0$ which implies $g = 0$.

If $j = 0$, then the proof is similar to the last step in item 3) above. □

Now we present the main result.

Theorem 5: The algebraic immunity of the function ϕ_{2k} obtained in Construction 2 equals k, for every $k \ge 0$.

Proof: We prove Theorem 5 by induction on k. There is nothing to check for $k = 0$. In the inductive step, we assume the hypothesis true until k and we have to prove that any nonzero function g_{2k+2} such that $g_{2k+2}\phi_{2k+2} = 0$ has degree at least $k+1$ (proving that any nonzero function g_{2k+2} such that $g_{2k+2}(\phi_{2k+2}+1) = 0$ has degree at least $k+1$ is similar). Suppose that such a function g_{2k+2} with degree at most k exists. Then, g_{2k+2} can be decomposed as

$$g_{2k+2} = g_{2k} \| g'_{2k} \| g''_{2k} \| h_{2k}$$

where $g_{2k}, g'_{2k}, g''_{2k} \in AN(\phi_{2k})$, and $h_{2k} \in AN(\phi_{2k}^1)$. The algebraic normal form of g_{2k+2} is then

$$\begin{aligned} g_{2k+2}(x_1, \ldots, x_{2k+2}) = {} & g_{2k} + x_{2k+1}(g_{2k} + g'_{2k}) \\ & + x_{2k+2}(g_{2k} + g''_{2k}) \\ & + x_{2k+1}x_{2k+2} \\ & \cdot (g_{2k} + g'_{2k} + g''_{2k} + h_{2k}). \end{aligned}$$

If g_{2k+2} has degree at most k, then $(g_{2k} + g'_{2k})$ and $(g_{2k} + g''_{2k})$ have degrees at most $k-1$. Because both functions lie in $AN(\phi_{2k})$ and $\mathcal{AI}_{2k}(\phi_{2k}) = k$, we deduce that $g_{2k} + g'_{2k} = 0$ and $g_{2k} + g''_{2k} = 0$, which give, $g_{2k} = g'_{2k} = g''_{2k}$. Therefore, $g_{2k+2} = g_{2k} + x_{2k+1}x_{2k+2}(g_{2k} + h_{2k})$

$$\deg(g_{2k}) \le k$$

and

$$\deg(g_{2k} + h_{2k}) \le k - 2.$$

According to Lemma 2, we have $g_{2k} = h_{2k}$. According to Lemma 3, we have then $g_{2k} = h_{2k} = 0$ that gives, $g_{2k+2} = 0$. This completes the proof. □

Remark 3: Let $f_l \in B_l$ be some l-variable function and let $f_{l+2k} = f_l + \phi_{2k}$, be the direct sum of f_l and ϕ_{2k}. Then we have the following results.

1) $nl(f_{l+2k}) = 2^l nl(\phi_{2k}) + 2^{2k} nl(f_l) - 2nl(\phi_{2k})nl(f_l) > 4^k nl(f_l)$.
2) If f_l is r-resilient, then f_{l+2k} is also r-resilient.
3) $\deg(f_{l+2k}) = \max\{\deg(f_l), \deg(\phi_{2k})\}$.
4) $\mathrm{wt}(f_{l+2k}) = \mathrm{wt}(\phi_{2k})(2^l - \mathrm{wt}(f_l)) + (2^{2k} - \mathrm{wt}(\phi_{2k}))\mathrm{wt}(f_l)$.
5) $\mathcal{AI}_{l+2k}(f_{l+2k}) \ge k+1$, for nonzero function f_l (see [30] for detailed proof).

In particular, if f_l is a nonconstant 1-variable function, we get a balanced function with optimum algebraic immunity.

Obviously, this is not the only secondary construction which can be used with ϕ_{2k}.

A. *Properties of the Constructed Functions*

1) Hamming Weight and Nonlinearity of ϕ_{2k}: We shall see that ϕ_{2k} is not balanced (it does not output as many 1's as 0's), but that its bias with respect to balancedness tends to zero when k tends to infinity.

Proposition 4: The Hamming weight of ϕ_{2k}^i

$$\mathrm{wt}\left(\phi_{2k}^i\right) = |\{x \in F_2^{2k} \mid \phi_{2k}^i(x) = 1\}|$$

equals $2^{2k-1} - \binom{2k-1}{k+i}$ for $k \geq 1$ and $i \geq 0$. Then

$$\mathrm{wt}(\phi_{2k}) = 2^{2k-1} - \binom{2k-1}{k}, \quad \text{for } k \geq 1.$$

Proof: Let us denote $w_{2k}^i = \mathrm{wt}\left(\phi_{2k}^i\right)$ and $w_{2k} = \mathrm{wt}(\phi_{2k}) = \mathrm{wt}\left(\phi_{2k}^0\right) = w_{2k}^0$. According to Relations (4) and (5), we can write: $w_0 = 0$

$$\begin{aligned} w_{2k} &= 3w_{2k-2} + w_{2k-2}^1, \quad \text{for any } k \geq 1 && (6)\\ w_{2k}^i &= w_{2k-2}^{i-1} + 2w_{2k-2}^i + w_{2k-2}^{i+1},\\ &\qquad \text{for any } k \geq 1, \quad i \geq 1 && (7) \end{aligned}$$

and

$$w_0^i = i[\mathrm{mod}\ 2], \qquad \text{for any } i > 0. \qquad (8)$$

Let us prove by induction on k that $w_{2k}^i = 2^{2k-1} - \binom{2k-1}{k+i}$, for any $k \geq 1$. This is true for $k = 1$ and for any $i \geq 0$. Indeed, we have

$$\begin{aligned} \phi_2^0 &= x_1x_2 = 0001\\ \phi_2^i &= (i-1)[\mathrm{mod}\ 2]\ (x_1+1)(x_2+1)\\ &\quad + i[\mathrm{mod}\ 2]\ (x_1+x_2) + (i+1)[\mathrm{mod}\ 2]\ x_1x_2\\ &= x_1 + x_2 + (i-1) \bmod 2, \quad \text{for any } i \geq 1. \end{aligned}$$

So, $w_2 = 1 = 2^1 - \binom{1}{1}$ and $w_2^i = 2 = 2^1 - \binom{1}{1+i}$ for $i > 0$.

We assume now that the induction assumption is true until $k-1$ and we prove it for k. Note that we have $2^{2k-1} = 3 \cdot 2^{2k-3} + 2^{2k-3}$ and $2^{2k-1} = 2^{2k-3} + 2 \cdot 2^{2k-3} + 2^{2k-3}$. We have also

$$\binom{2k-1}{k} = 3 \cdot \binom{2k-3}{k-1} + \binom{2k-3}{k}$$

since

$$\begin{aligned} &3 \cdot \binom{2k-3}{k-1} + \binom{2k-3}{k}\\ &= \frac{(2k-3)!}{k!(k-2)!}(3k+k-2) = \frac{(2k-3)!}{k!(k-2)!}(4k-2) = \frac{(2k-1)!}{k!(k-1)!} \end{aligned}$$

and

$$\begin{aligned} \binom{2k-1}{k+i} &= \binom{2k-2}{k+i-1} + \binom{2k-2}{k+i}\\ &= \binom{2k-3}{k+i-2} + 2 \cdot \binom{2k-3}{k+i-1} + \binom{2k-3}{k+i}. \end{aligned}$$

So, (6) and (7) and the induction assumption imply that

$$w_{2k}^i = 2^{2k-1} - \binom{2k-1}{k+i}. \qquad \square$$

The weight $2^{2k-1} - \binom{2k-1}{k-1})$ of ϕ_{2k} is therefore equivalent to $2^{2k-1}(1 - \frac{1}{\sqrt{\pi k}})$, which is asymptotically balanced.

Given the recent result of [36], we can exactly calculate the nonlinearity.

Theorem 6: $nl(\phi_{2k}) = 2^{2k-1} - \binom{2k-1}{k-1}$.

Proof: Consider the $(2k+1)$-variable function $x_{2k+1} + \phi_{2k}(x_1, \ldots, x_{2k})$. This function is at a distance

$$2\left(2^{2k-1} - \binom{2k-1}{k-1}\right) = 2^{2k} - \binom{2k}{k}$$

from the linear function x_{2k+1}. Thus, $nl(x_{2k+1} + \phi_{2k}) \leq 2^{2k} - \binom{2k}{k}$. Since $x_{2k+1} + \phi_{2k}$ is of full algebraic immunity $k+1$ [30], following [36, Corollary 1], one gets $nl(x_{2k+1} + \phi_{2k}) \geq 2^{2k} - \binom{2k}{k}$. Thus,

$$nl(x_{2k+1} + \phi_{2k}) = 2^{2k} - \binom{2k}{k}. \qquad (9)$$

It is well known and easily checked that, for every $2k$-variable function f, we have $nl(x_{2k+1} + f) = 2nl(f)$. This completes the proof. $\square$

Note that $nl(\phi_{2k}) = 2^{2k-1} - \binom{2k-1}{k-1}$, which is strictly greater than the lower bound $2^{2k-1} - \binom{2k}{k}$ as presented in [36, Corollary 2]. However, the nonlinearity of the function ϕ_{2k} is not very good. The ratio $\frac{2^{2k-1} - nl(\phi_{2k})}{2^{2k-1} - nlmax}$, where $nlmax = 2^{2k-1} - 2^{k-1}$ is the maximum possible nonlinearity of Boolean functions, equals $\frac{\binom{2k-1}{k-1}}{2^{k-1}} \sim \frac{2^k}{\sqrt{k\pi}}$ and is therefore not sufficient for use of the function without using a secondary construction to enhance the nonlinearity.

2) Algebraic Degree of ϕ_{2k}: When the weight of the function ϕ_{2k} is odd, then clearly its algebraic degree is $2k$. We shall subsequently prove that, when the weight is even, the algebraic degree is also very high. Note that, denoting respectively by c_{2k} and c_{2k}^i the 2-variable functions equal to the factors of $\prod_{j=3}^{2k} x_j$ in the ANFs of ϕ_{2k} and ϕ_{2k}^i, (4) and (5) straightforwardly imply that $c_{2k+2} = c_{2k} + c_{2k}^1$ and $c_{2j}^i = c_{2j-2}^{i-1} + c_{2j-2}^{i+1}$. But it seems difficult to find directly the exact expression of c_{2k} satisfying these constraints (with the initialization $c_{2j}^0 = c_{2j}$ for $j \geq 0$, $c_0^i = i\ [\mathrm{mod}\ 2]$ for $i \geq 0$). We shall observe that, changing the initialization in the recursive definition of c_{2k}, and using only (5) in the recursion leads to an affine function q_{2k}. Considering then the difference between q_{2k} and c_{2k} will permit to obtain a recurrence relation on c_{2k} which will not involve the c_{2k}^i's. In fact, in the next proof, we shall be able to partially develop this method directly on ϕ_{2k}, which will give more information; we will deduce then our result on c_{2k} through a simple reduction.

Proposition 5: Let c_{2k} be the 2-variable function equal to the factor of $\prod_{j=3}^{2k} x_j$ in the ANF of ϕ_{2k}. Then we have

$$\begin{aligned} c_{2k} &= (\lfloor(\log_2(k)\rfloor\ [\mathrm{mod}\ 2]\ )\,(x_1 + x_2)\\ &\qquad + \sum_{0 \leq t \leq \log_2(k)} c_{2(k-2^t)} + cst \qquad (10) \end{aligned}$$

where cst is some bit depending on k.

Proof: Let q_{2j}^i be recursively defined by (5), with the initialization $q_0^i = i[\mathrm{mod}\ 2]$, for any positive or negative i. We have $q_{2j}^i = x_1 + x_2 + \cdots + x_{2j-1} + x_{2j} + (i+j)[\ \mathrm{mod}\ 2]$ (this can be easily checked by induction). The function $r_{2j}^i = \phi_{2j}^i + q_{2j}^i$ satisfies (5) and we have $r_0^i = 0$ for $i > 0$, $r_{2j}^0 = \phi_{2j} + q_{2j}^0$ for $j \geq 0$.

The function r_{2j}^i is the addition of what we collect with (5) (or its translation in terms of ANF) for all the paths starting from the point of coordinates $(i, 2j)$ and arriving for the first time to a point of coordinates $(0, 2l)$ (where $2l$ can be any integer between 0 and $2(j-i)$), and constituted of the concatenation of elementary paths going from a point of coordinates $(s, 2t)$ to a point of coordinates $(s+\epsilon, 2t-2)$, $\epsilon \in \{-1, 0, 1\}$.

We deduce that

$$r^1_{2k-2} = \sum_{l=0}^{k-2} \left(\left(\phi_{2l} + q^0_{2l}\right) \sum_{\substack{\epsilon\in\{-1,0,1\}^{k-1-l};\\ \epsilon_1+\cdots+\epsilon_{k-1-l}=-1;\\ \forall t>1, \epsilon_t+\cdots+\epsilon_{k-1-l}\geq 0}} \prod_{r=1}^{k-1-l} U_{\epsilon_r, l+r} \right)$$

where $U_{-1,r} = (x_{2r-1}+1)(x_{2r}+1)$, $U_{0,r} = x_{2r-1}+x_{2r}$, and $U_{1,r} = x_{2r-1}x_{2r}$.

This implies, for $k \geq 3$

$$c^1_{2k-2} = \sum_{l=0}^{k-2} \left((c_{2l} + x_1 + x_2 + cst) \sum_{\substack{\epsilon\in\{-1,1\}^{k-1-l};\\ \epsilon_1+\cdots+\epsilon_{k-1-l}=-1;\\ \forall t>1, \epsilon_t+\cdots+\epsilon_{k-1-l}\geq 0}} 1 \right)$$

where the cst's are constants. The number

$$\eta_l = \sum_{\substack{\epsilon\in\{-1,1\}^{k-1-l};\\ \epsilon_1+\cdots+\epsilon_{k-1-l}=-1;\\ \forall t>1, \epsilon_t+\cdots+\epsilon_{k-1-l}\geq 0}} 1$$

equals 1 for $l = k-2$; it equals 0 for $l = k-3$; and for $l \leq k-4$, it equals the number of paths from the point of coordinates $(i, 2j)$ where $i = 2$; $j = k-2$ to the point of coordinates $(i, 2j)$ where $i = 2$; $j = l+2$ which do not cut the axis of equation $i = 0$ (indeed, the two last elementary paths are necessarily $(2, 2l+4) \rightarrow (1, 2l+2)$ and $(1, 2l+2) \rightarrow (0, 2l)$). Note that η_l is null if $k-l$ is odd. We assume now that $k-l$ is even. Then η_l equals the number of all paths between these two points (the points $(2, 2k-4)$ and $(2, 2l+4)$), that is, $\binom{k-l-4}{\frac{k-l}{2}-2}$, minus the number of paths cutting the axis $i = 0$. This last number equals the number of paths from the point of coordinates $(i, 2j)$ where $i = 2$; $j = k-2$ to the point of coordinates $(i, 2j)$ where $(i = -2; j = l+2)$ (replacing the lower part of each path cutting the axis by its mirror image with respect to this axis). We have then

$$\eta_l = \binom{k-l-4}{\frac{k-l-4}{2}} - \binom{k-l-4}{\frac{k-l-4}{2}+2}.$$

Hence, since $\binom{k-l-4}{\frac{k-l-4}{2}}$ is even for $k-l$ even greater than 4 and equals 1 for $k-l = 4$, and according to Lucas' theorem, $\eta_l[\bmod 2]$ equals 1 if and only if $k-l$ is a power of 2. We deduce, denoting $k-l$ by 2^t

$$c^1_{2k-2} = \sum_{1\leq t\leq \log_2(k)} (c_{2(k-2^t)} + x_1 + x_2) + cst.$$

This completes the proof. □

For small indices, this gives

$$\begin{aligned}
c_2 &= x_1x_2\\
c_4 &= x_1 + x_2 + c_2 + c_0 + cst = x_1x_2 + x_1 + x_2 + cst\\
c_6 &= x_1 + x_2 + c_4 + c_2 + cst = cst,\\
c_8 &= c_6 + c_4 + c_0 + cst = x_1x_2 + x_1 + x_2 + cst,\\
c_{10} &= c_8 + c_6 + c_2 + cst = x_1 + x_2 + cst\\
c_{12} &= c_{10} + c_8 + c_4 + cst = x_1 + x_2 + cst,\\
c_{14} &= c_{12} + c_{10} + c_6 + cst = cst\\
c_{16} &= x_1 + x_2 + c_{14} + c_{12} + c_8 + c_0 + cst\\
&= x_1x_2 + x_1 + x_2 + cst\\
c_{18} &= x_1 + x_2 + c_{16} + c_{14} + c_{10} + c_2 + cst\\
&= x_1 + x_2 + cst\\
c_{20} &= x_1 + x_2 + c_{18} + c_{16} + c_{12} + c_4 + cst\\
&= x_1 + x_2 + cst\\
c_{22} &= x_1 + x_2 + c_{20} + c_{18} + c_{14} + c_6 + cst\\
&= x_1 + x_2 + cst.
\end{aligned}$$

The coefficient v_{2k} of x_1x_2 in c_{2k} satisfies the relation $v_{2k} = \sum_{0\leq t\leq \log_2(k)} v_{2(k-2^i)}$ and $v_0 = 0$, $v_2 = 1$. It can be shown by induction that $v_{2k} = 1$, i.e., ϕ_{2k} has degree $2k$, if and only if k is a power of 2 (but we knew this already thanks to Lucas' theorem and to Proposition 4). Similarly, the coefficient of x_1 (or, of x_2) in c_{2k} equals 0 if and only if $k+1$ is a power of 2. Hence, ϕ_{2k} has most often degree exactly $2k-1$. We deduce the following.

Proposition 6: For $k \geq 1$ the degree of ϕ_{2k} is as follows.

1) $\deg(\phi_{2k}) = 2k$ if and only if k is a power of 2.
2) If neither k nor $k+1$ is a power of 2, then $\deg(\phi_{2k}) = 2k-1$.
3) If $k+1$ is a power of 2, then $2k-3 \leq \deg(\phi_{2k}) \leq 2k-1$.

Proof: From the preceding discussion, items 1) and 2) are proved. For item 3), if $k+1$ is a power of 2, then $\deg(\phi_{2k}) \leq 2k-1$. Since

$$\phi_{2k} = \phi_{2k-2} + x_{2k-1}x_{2k}\left(\phi_{2k-2} + \phi^1_{2k-2}\right)$$

we have $\deg(\phi_{2k}) \geq \deg(\phi_{2k-2})$. So, for item 3), $\deg(\phi_{2k}) \geq 2k-3$. □

3) The Structure of ϕ_{2k} and an Efficient Way of Computing its Output: Here we study the structure of ϕ_{2k}. We observe that the function ϕ_{2k} can be written as the sum of two functions ϕ'_{2k} and ϕ''_{2k}, which can be obtained from symmetric functions by the same transformation easy to implement. Let ${\phi'}^i_{2k}$ and ${\phi''}^i_{2k}$ be the sequences of Boolean functions satisfying (4) and (5) for every $k \geq 2$, and initialized as follows: ${\phi'}^0_2 = x_1x_2+x_1+x_2+1$ (that is, ${\phi'}^0_2 = \delta_0(x_1, x_2)$, the indicator of $\{0, 0\}$) and ${\phi'}^i_2 = 0$ for $i \geq 1$, ${\phi''}^i_2 = x_1+x_2+i+1\,[\bmod\ 2]$ for $i \geq 0$. The function ϕ_{2k} equals the sum of ${\phi'}^0_{2k}$ and ${\phi''}^0_{2k}$, since the sum of ${\phi'}^i_{2k}$ and ${\phi''}^i_{2k}$ equals ϕ^i_{2k} for $k = 1$ and $i \geq 0$ and satisfies the same recurrence relations. Note that, since ${\phi'}^0_2$ equals δ_0 and ${\phi'}^i_2 = 0$, for every $i > 1$, the restriction of ${\phi'}^0_{2k}$ to the set of words whose two first coordinates are not both null is constantly equal to zero. We shall see that the restriction of ${\phi'}^0_{2k}$ to the set of words whose two first coordinates are both null—let us denote this function by ${\psi'}^0_{2k-2}$—is related to the Boolean function on F_2^{2k-2} equal to the indicator of the set of vectors whose Hamming weights equal $k-1$ and $k-2$.

Proposition 7: Let ${\varphi'}^i_{2k}$ be the $2k$-variable Boolean function recursively defined by the following relations:

- for any $k \geq 1$, ${\varphi'}^0_{2k} = \overline{{\varphi'}^0_{2k-2}}||{\varphi'}^0_{2k-2}||{\varphi'}^0_{2k-2}||{\varphi'}^1_{2k-2}$, where the truth table of $\overline{{\varphi'}^0_{2k-2}}$ is the reverse of the truth table of ${\varphi'}^0_{2k-2}$ (hence, $\overline{{\varphi'}^0_{2k-2}}(x_1,\ldots,x_{2k-2}) = {\varphi'}^0_{2k-2}(x_1+1,\ldots,x_{2k-2}+1)$); in terms of ANF
$$\begin{aligned}{\varphi'}^0_{2k} = (x_{2k-1}+1)(x_{2k}+1)\overline{{\varphi'}^0_{2k-2}} + (x_{2k-1}\\ +x_{2k}){\varphi'}^0_{2k-2} + x_{2k-1}x_{2k}{\varphi'}^1_{2k-2};\end{aligned}$$
- for any $k \geq 1$ and any $i \geq 1$
$${\varphi'}^i_{2k} = {\varphi'}^{i-1}_{2k-2}||{\varphi'}^i_{2k-2}||{\varphi'}^i_{2k-2}||{\varphi'}^{i+1}_{2k-2};$$
in terms of ANF
$$\begin{aligned}{\varphi'}^i_{2k} = (x_{2k-1}+1)(x_{2k}+1){\varphi'}^{i-1}_{2k-2} + (x_{2k-1}\\ +x_{2k}){\varphi'}^i_{2k-2} + x_{2k-1}x_{2k}{\varphi'}^{i+1}_{2k-2};\end{aligned}$$
- ${\varphi'}^0_0 = 1$ and ${\varphi'}^i_0 = 0$ for any $i \geq 1$.

Then, for any $k \geq 0$ and any $i \geq 0$, ${\varphi'}^i_{2k}$ equals the indicator of the set of vectors of F_2^{2k} whose Hamming weights equal $k-i$ and $k-i-1$.

Proof: We show this by induction on k. It is true for $k=0$. Let us prove that, if it is true for some $k \geq 0$, then it is true for $k+1$.

— if $i \geq 1$, a vector $(x_1,\ldots,x_{2k+2})$ whose Hamming weight equals $k+1-i$ or $k-i$ either equals $x00$, where x is a vector of length $2k$ whose Hamming weight equals $k+1-i = k-(i-1)$ or $k-i = k-(i-1)-1$, or equals $x10$ or $x01$, where x has Hamming weight $k-i$ or $k-i-1$, or equals $x11$, where x has Hamming weight $k-i-1 = k-(i+1)$ or $k-i-2 = k-(i+1)-1$. This corresponds to the relation ${\varphi'}^i_{2k+2} = {\varphi'}^{i-1}_{2k}||{\varphi'}^i_{2k}||{\varphi'}^i_{2k}||{\varphi'}^{i+1}_{2k}$;

— if $i=0$, a vector $(x_1,\ldots,x_{2k+2})$ whose Hamming weight equals $k+1$ or k either equals $(x+\mathbf{1})00$, where $\mathbf{1}$ is the all-one vector of length $2k$ and x has Hamming weight $2k-(k+1) = k-1$ or $2k-k = k$, or equals $x10$ or $x01$, where x has Hamming weight k or $k-1$, or equals $x11$, where x has Hamming weight $k-1$ or $k-2$. This corresponds to the relation ${\varphi'}^0_{2k+2} = \overline{{\varphi'}^0_{2k}}||{\varphi'}^0_{2k}||{\varphi'}^0_{2k}||{\varphi'}^1_{2k}$. □

Remark 4: Note that the ANF of ${\varphi'}^0_{2k}$ is easy to obtain, since ${\varphi'}^0_{2k}$ is symmetric: for every $I \subseteq \{1,\ldots,2k\}$, the coefficient of the monomial $\prod_{j\in I} x_j$ in the ANF of ${\varphi'}^0_{2k}$ is $\binom{|I|}{k} + \binom{|I|}{k-1}$ [mod 2], and Lucas' theorem gives then its effective value. More importantly, the output to ${\varphi'}^0_{2k}$ is quite easy to compute, with less than $2k$ additions with carries.

Let us see what transformation on the truth tables of the functions permits, from the value of ${\varphi'}^0_{2k-2}$, to obtain the value of ${\psi'}^0_{2k-2}$ (i.e., of the restriction of ${\phi'}^0_{2k}$ to the set of those words whose two first coordinates are both null—we know that, for the other vectors, ${\phi'}^0_{2k}$ takes null value). The only difference between ${\psi'}^0_{2k}$ and ${\varphi'}^0_{2k}$ is in the relations
$${\psi'}^0_{2k+2} = {\psi'}^0_{2k}||{\psi'}^0_{2k}||{\psi'}^0_{2k}||{\psi'}^1_{2k}$$
and
$${\varphi'}^0_{2k+2} = \overline{{\varphi'}^0_{2k}}||{\varphi'}^0_{2k}||{\varphi'}^0_{2k}||{\varphi'}^1_{2k}.$$

When calculating recursively the value of the function ${\varphi'}^0_{2k}$ at a vector $(x_1,\ldots,x_{2k})$, we arrive to a situation where ${\varphi'}^0_{2l}$ must be reversed when, reading $x = (x_1,\ldots,x_{2k})$ from its rightmost position to its leftmost position, the number of times we encountered $x_{2l} = x_{2l-1} = 1$, minus the number of times we encountered $x_{2l} = x_{2l-1} = 0$ equals -1, for the first time. We then have to complement all of the remaining coordinates of x and apply recursively the same transformation to these remaining coordinates.

For instance, for $x = 000100$, we start with a null difference (between the numbers of 11 and of 00), we read 00, so the difference is -1, and we have to complement 0001 into 1110; after reading the two rightmost bits of this last vector, the difference remains null and after reading the two leftmost bits, it is 1. So we have ${\varphi'}^0_6(000100) = {\psi'}^0_6(111000)$. Let us give a few other examples:
$$\begin{aligned}{\varphi'}^0_6(001100) &= {\psi'}^0_6(000000)\\ {\varphi'}^0_6(100000) &= {\psi'}^0_6(011100)\\ {\varphi'}^0_6(010010) &= {\psi'}^0_6(100010)\end{aligned}$$
and
$${\varphi'}^0_6(100001) = {\psi'}^0_6(010001).$$

The inverse of this transformation is easy to deduce: we apply a similar principle, but each time we have applied a complementation according to the count on the previous pairs of bits, we complement according to the count of previous pairs of bits of the input vector. This gives the following algorithm, the transformation from $(x_1,\ldots,x_{2k})$ to $(y_1,\ldots,y_{2k})$ such that $\psi'(x_1,\ldots,x_{2k}) = \varphi'(y_1,\ldots,y_{2k})$.

Algorithm 1: Input $x = (x_1,\ldots,x_{2k})$. Output $y = (y_1\ldots,y_{2k})$

```
• Initialize $y$ by $x$;
• count := 0; $i : 2k - 1$;
• while $i \geq 0$ do
    —if $x_i = x_{i+1} = 1$ then count := count + 1;
    —else if $x_i = x_{i+1} = 0$ then
        ⋆ count := count − 1;
        ⋆ If count := −1 then replace $(y_1,\ldots,y_{i-1})$
            by $(y_1 + 1,\ldots,y_{i-1} + 1)$ and apply count := 0.
    —$i := i - 2$;
```

The output to this algorithm is the transformation (inverse of the above discussed transformation) from $(x_1,\ldots,x_{2k})$ to $(y_1,\ldots,y_{2k})$ such that $\psi'(x_1,\ldots,x_{2k}) = \varphi'(y_1,\ldots,y_{2k})$. So, to compute $\psi'(x_1,\ldots,x_{2k})$, first we run Algorithm 1 to get the transformation $(y_1,\ldots,y_{2k})$. Then $\varphi'(y_1,\ldots,y_{2k})$ computes the value for $\psi'(x_1,\dots,x_{2k})$. Hence, we have here a very fast way of computing the output to ψ'_{2k}, and therefore to ϕ'_{2k}.

Let us see now that the situation with ϕ''_{2k} is similar. By definition, for any k and any i, ϕ''_{2k} is the concatenation of functions equal to x_1+x_2 and x_1+x_2+1. Hence, it is the direct sum of x_1+x_2 and of the function ${\psi''}^{i}_{2k-2}$, where ${\psi''}^{i}_{2k}$ is the $2k$-variable function such that ${\psi''}^{0}_{2k}$ equals

$${\psi''}^0_{2k-2}\|{\psi''}^0_{2k-2}\|{\psi''}^0_{2k-2}\|{\psi''}^1_{2k-2}$$

for every $k\geq 1$, ${\psi''}^i_{2k}$ equals

$${\psi''}^{i-1}_{2k-2}\|{\psi''}^{i}_{2k-2}\|{\psi''}^{i}_{2k-2}\|{\psi''}^{i+1}_{2k-2}$$

for every $k\geq 1$ and every $i\geq 1$ and ${\psi''}^i_0=i+1$ [mod 2].

Proposition 8: Let ${\varphi''}^i_{2k}$ be the $2k$-variable Boolean function recursively defined by the following relations:

- for any $k\geq 1$
$${\varphi''}^0_{2k}=\overline{{\varphi''}^0_{2k-2}}\|{\varphi''}^0_{2k-2}\|{\varphi''}^0_{2k-2}\|{\varphi''}^1_{2k-2};$$
- for any $k\geq 1$ and any $i\geq 1$
$${\varphi''}^i_{2k}={\varphi''}^{i-1}_{2k-2}\|{\varphi''}^{i}_{2k-2}\|{\varphi''}^{i}_{2k-2}\|{\varphi''}^{i+1}_{2k-2};$$
- ${\varphi''}^i_0=i+1$ [mod 2].

Then, for any $k\geq 0$ and any $i\geq 0$, ${\varphi''}^i_{2k}$ equals the indicator of the set of vectors of F_2^{2k} whose Hamming weights belong to one of the two sets

$$\{j\in\{0,\dots,k-i-1\}\mid j\equiv k-i-1 \bmod 2\};$$
$$\{j\in\{k-i,\dots,2k\}\mid j\equiv k-i \bmod 2\}.$$

Proof: Let us show this by induction. This is true for $k=0$. If it is true for some $k\geq 0$ then it is true for $k+1$

— if $i\geq 1$, a vector $(x_1,\dots,x_{2k+2})$ whose Hamming weight belongs to

$$\{j\in\{0,\dots,(k+1)-i-1\}\mid j\equiv (k+1)-i-1 \bmod 2\}$$
$$\cup\{j\in\{(k+1)-i,\dots,2k\}\mid j\equiv (k+1)-i \bmod 2\}$$

either equals $x00$, where x is a vector of length $2k$ whose Hamming weight belongs to this same set, that is, to

$$\{j\in\{0,\dots,k-(i-1)-1\}\mid j\equiv k-(i-1)-1 \bmod 2\}$$
$$\cup\{j\in\{k-(i-1),\dots,2k\}\mid j\equiv k-(i-1) \bmod 2\}$$

or equals $x10$ or $x01$, where the Hamming weight of x belongs to

$$\{j\in\{0,\dots,k-i-1\}\mid j\equiv k-i-1 \bmod 2\}$$
$$\cup\{j\in\{k-i,\dots,2k\}\mid j\equiv k-i \bmod 2\}$$

or equals $x11$, where the Hamming weight of x belongs to

$$\{j\in\{0,\dots,k-(i+1)-1\}\mid j\equiv k-(i+1)-1 \bmod 2\}$$
$$\cup\{j\in\{k-(i+1),\dots,2k\}\mid j\equiv k-(i+1) \bmod 2\}.$$

This corresponds to the relation

$${\varphi''}^i_{2k+2}={\varphi''}^{i-1}_{2k}\|{\varphi''}^{i}_{2k}\|{\varphi''}^{i}_{2k}\|{\varphi''}^{i+1}_{2k}.$$

— if $i=0$, a vector $(x_1,\dots,x_{2k+2})$ whose Hamming weight belongs to

$$\{j\in\{0,\dots,(k+1)-1\}\mid j\equiv (k+1)-1 \bmod 2\}$$
$$\cup\{j\in\{(k+1),\dots,2k\}\mid j\equiv (k+1) \bmod 2\}$$

either equals $(x+\mathbf{1})00$, where the Hamming weight of x belongs to

$$\{j\in\{k,\dots,2k\}\mid j\equiv k \bmod 2\}$$
$$\cup\{j\in\{0,\dots,k-1\}\mid j\equiv k-1 \bmod 2\}$$

or equals $x10$ or $x01$, where the Hamming weight of x belongs to

$$\{j\in\{0,\dots,k-1\}\mid j\equiv k-1 \bmod 2\}$$
$$\cup\{j\in\{k,\dots,2k\}\mid j\equiv k \bmod 2\}$$

or equals $x11$, where the Hamming weight of x belongs to

$$\{j\in\{0,\dots,k-2\}\mid j\equiv k-2 \bmod 2\}$$
$$\cup\{j\in\{k-1,\dots,2k\}\mid j\equiv k-1 \bmod 2\}.$$

This corresponds to the relation

$${\varphi''}^0_{2k+2}=\overline{{\varphi''}^0_{2k}}\|{\varphi''}^0_{2k}\|{\varphi''}^0_{2k}\|{\varphi''}^1_{2k}.\qquad\square$$

Function ψ''_{2k} can be obtained from φ''_{2k} by the same algorithm as the one giving ψ'_{2k} from φ'_{2k}.

We now present the final algorithm to compute $\phi_{2k}(x_1,\dots,x_{2k})$. To compute this we need a preprocessing step to establish two symmetric functions φ' and φ'' on $(2k-2)$ variables. As the two functions are symmetric, the following preprocessing step calculates the short truth tables (corresponding to each input weight) of those two functions. Here we use the notation for the output to a symmetric function s at inputs of weight i as $s[i]$.

Preprocessing:

```
• for ($i=0$; $i\leq 2k-2$; $i++$)
    — $\varphi'[i]=\varphi''[i]=0$;
• $\varphi'[k-2]=\varphi'[k-1]=1$;
• for ($i=0$; $i\leq k-2$; $i++$)
    — if ($i\equiv k-2 \bmod 2$) then $\varphi''[i]=1$
• for ($i=k-1$; $i\leq 2k-2$; $i++$)
    — if ($i\equiv k-1 \bmod 2$) then $\varphi''[i]=1$
```

Now the following algorithm calculates $\phi(x)$.

Algorithm 2: Input: $x=(x_1,\dots,x_{2k})$.

```
• $y=(y_1,\dots,y_{2k-2})\leftarrow(x_3,\dots,x_{2k})$.
• Apply Algorithm 1 on the $2k-2$ bits vector $y$.
• $w\leftarrow \mathrm{wt}(y)$.
• Output: $(1+(x_1\vee x_2))\varphi'[w]+(x_1+x_2)+\varphi''[w]$.
```

So, following Algorithm 2, the output to ϕ_{2k} can be very efficiently computed. Precisely, the number of elementary operations which have to be performed for calculating the output to ϕ_{2k} is less than $12k$, since, for each of the functions ϕ'_{2k} and ϕ''_{2k}, it equals the number of complementations and additions (with carries) to apply Algorithm 1, that is, at most $4k$, plus the number of additions with carries to calculate the output to a $2k-2$-variable symmetric function, plus, in the case of ψ''_{2k}, the addition (without carry) of a 2-variable affine function. This means that we can use ϕ_{2k} in a stream cipher, with the same efficiency as when we used a function defined by a lookup table or by an ANF, with $\log_2(12k)$ variables (in practice, we shall have to choose ϕ_{2k} with fewer variables and to use it in a secondary construction, to obtain a balanced—and if necessary resilient—function with a good nonlinearity). For instance, if $\log_2(12k)$ equals 8, then $2k$ equals 44; if $\log_2(12k)$ equals 9, then $2k$ equals 86, and if $\log_2(12k)$ equals 10, then $2k$ equals 172. Recall that, before the existence of algebraic attacks, Boolean functions used in stream ciphers had usually at most ten variables, for reasons of efficiency (unless they were peculiar as in the case of LILI-128, but their peculiarities have been responsible of their weaknesses with respect to algebraic attacks).

Thanks to the observations above, it may be possible to deduce the ANF's of ϕ'_{2k}, ϕ''_{2k} from those of $\varphi'_{2k}, \varphi''_{2k}$. But the fact that we could obtain a fast way of computing the output to ϕ_{2k} is more practically interesting than obtaining its ANF.

B. *Different Initializations in Construction 2*

A drawback of the function ϕ_{2k} is that it is unbalanced. This happens since $\phi_2 = x_1x_2$ is unbalanced. If one starts the construction with ϕ_2 as affine function, then the function ϕ_{2k} will always be balanced as $\phi^i_{2j} = x_1+x_2+(i+j) \bmod 2$ for $i>0$, $j>0$. Now we present some observations in this regard.

1) Take $\phi_2 = x_1 + x_2$.

Case 1: $\phi^i_2 = x_1 + x_2$ if i is even and $\phi^i_2 = 1 + x_1 + x_2$ if i is odd for $i > 0$. These are presented in the following table.

Case 2: Also in brackets, we present the results when $\phi^i_2 = x_1 + x_2$ if i is odd and $\phi^i_2 = 1 + x_1 + x_2$ if i is even for $i > 0$.

function	degree	nonlinearity	resiliency	$\mathcal{AI}$
ϕ_2	1(1)	0(0)	1(1)	1(1)
ϕ_4	2(1)	4(0)	1(1)	2(1)
ϕ_8	4(4)	20(0)	1(1)	3(2)
ϕ_8	5(6)	88(28)	1(1)	4(3)
ϕ_{10}	8(8)	372(148)	1(1)	5(4)

Here, for the first case, ϕ_{2k} is always 1-resilient, optimal algebraic immunity is achieved and nonlinearity is slightly lesser than what we have observed for Construction 2. However, in the second case, ϕ_{2k} has poor nonlinearity and lower AI.

2) Then we have attempted $\phi_2 = x_1$ and $\phi^i_2 = x_1 + x_2$ when i is even (respectively, odd) and $\phi^i_2 = 1+x_1+x_2$ when i is odd (respectively, even). We found algebraic immunity is optimal but poor nonlinearity. The results are same for both the cases so we do not write them separately in brackets.

function	degree	nonlinearity	resiliency	$\mathcal{AI}$
ϕ_2	1	0	0	1
ϕ_4	3	2	0	2
ϕ_6	4	12	0	3
ϕ_8	7	58	0	4
ϕ_{10}	8	260	0	5

3) Take $\phi_2 = x_1$ and $\phi^i_2 = x_1 + x_2$, $i > 0$. We find that the ANF of ϕ_{2k} is of the form $\phi_{2k} = x_1 + x_2F$, where F is a function on $2k-2$ many variables. So, AI will be ≤ 2, since $(1+x_1)(1+x_2)$ is annihilator of ϕ_{2k} for any $k > 1$.

So it seems that just by changing the initializations in Construction 2, it may not be possible to get dramatically better results. One may need to attempt for completely different kinds of construction to achieve better parameters.

VI. Conclusion

In this paper, the algebraic immunity property of a Boolean function was studied in great details. We first identified a fundamental relationship between the Walsh spectrum and algebraic immunity of a Boolean function, leading to a lower bound on the nonlinearity (during the review process of our paper, an improvement of this bound has been found by Lobanov [36]); this question of knowing whether these two criteria were opposite or not was challenging and this shows that they are not. Moreover, we showed similar relationship with the higher order nonlinearities. We followed with certain enumeration results of independent annihilators, which have some interest from computational viewpoint. Then we have studied some existing constructions in terms of their algebraic immunity, both theoretically and experimentally; this was necessary for practical design of cryptographic functions. Next, we presented a construction of Boolean functions with maximum possible algebraic immunity and studied the cryptographic properties of the construction. The constructed functions have high degrees but are not balanced and have insufficient nonlinearity. However, they can be used in secondary constructions settling these drawbacks. We gave an algorithm to compute their outputs, which makes them as easy to calculate as all the other known (infinite classes of) functions with best possible algebraic immunity. All of these other functions are symmetric and present therefore a risk if attacks using this peculiarity can be found in the future. We believe this makes the functions studied here more interesting for a use in stream ciphers.

The field is still open in many aspects. To be specific, getting a primary construction with optimum properties in terms of algebraic immunity and several other cryptographic properties (balancedness, nonlinearity), and avoiding dangerous peculiarities, looks extremely challenging at this point of time, since it may provide a more efficient design of cryptographic functions meeting all the necessary criteria for being used in stream ciphers (however, the question of a fast implementation of the functions will have to be also addressed). In this consideration, the method of construction presented in this paper opens a completely new way of designing cryptographic functions having provably optimum algebraic immunity and which can be very efficiently implemented.

Acknowledgment

The authors wish to thank the anonymous reviewers for their comments that improved the technical quality and the editorial presentation of this paper. They ualso thank the anonymous reviewers of Indocrypt 2004 and FSE 2005 for their excellent comments over the conference versions. We acknowledge Prof. Yuriy Tarannikov for pointing out some gaps in the proof of maximum algebraic immunity that has been published in FSE 2005 (the proof was also written in a complicated manner). The discussion with Prof. Tarannikov helped in getting a more concrete and clearer proof.

References

[1] F. Armknecht, C. Carlet, P. Gaborit, S. Kuenzli, W. Meier, and O. Ruatta, "Efficient computation of algebraic immunity for algebraic and fast algebraic attacks," in *Proc. EUROCRYPT 2006 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2006, vol. 4004, pp. 147–164.

[2] F. Armknecht and M. Krause, "Algebraic attacks on combiners with memory," in *Advances in Cryptology—CRYPTO 2003 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2003, vol. 2729, pp. 162–175.

[3] F. Armknecht, "Improving fast algebraic attacks," in *Proc. Workshop on Fast Software Encryption (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2004, vol. 3017, pp. 65–82.

[4] L. M. Batten, "Algebraic attacks over $\mathrm{GF}(q)$," in *Progress in Cryptology—INDOCRYPT 2004 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, , 2004, vol. 3348, pp. 84–91.

[5] A. Botev, "On algebraic immunity of some recursively given sequence of correlation immune functions," in *Proceedings of XV International Workshop on Synthesis and Complexity of Control Systems* (in Russian), Novosibirsk, Oct. 18-23, 2004, pp. 8–12.

[6] ——, "On algebraic immunity of new constructions of filters with high nonlinearity," in *Proce. VI Int. Conf. Discrete Models in the Theory of Control Systems* (in Russian), Moscow, Russia, Dec. 2004, pp. 227–230.

[7] A. Botev and Y. Tarannikov, "Lower bounds on algebraic immunity for recursive constructions of nonlinear filters," Preprint, 2004.

[8] A. Braeken, J. Lano, N. Mentens, B. Preneel, and I. Verbauwhede, "SFINKS: A synchronous stream cipher for restricted hardware environments," in *Proc. SKEW—Symmetric Key Encryption Workshop*, Aarhus, Denmark, May 2005.

[9] A. Braeken, J. Lano, and B. Preneel, "Evaluating the resistance of filters and combiners against fast algebraic attacks," Eprint on ECRYPT, 2005.

[10] A. Braeken and B. Preneel, "On the algebraic immunity of symmetric Boolean functions," in *Indocrypt 2005 (Lecture Notes in Computer Science)*, Jul. 26, 2005, vol. 3797, pp. 35–48 [Online]. Available: http://eprint.iacr.org/

[11] P. Camion, C. Carlet, P. Charpin, and N. Sendrier, "On correlation-immune functions," in *Advances in Cryptology: Crypto '91, Proceedings (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1991, vol. 576, pp. 86–100.

[12] A. Canteaut and M. Trabbia, "Improved fast correlation attacks using parity-check equations of weight 4 and 5," in *EUROCRYPT 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1807, pp. 573–588.

[13] C. Carlet, "Recent results on binary bent functions," *J. Comb., Inf., System Sci. (Proc. Int. Conf. Combinatorics, Information Theory and Statistics)*, vol. 25, pp. 133–149, 2000.

[14] C. Carlet, "A larger class of cryptographic Boolean functions via a study of the Maiorana-McFarland construction," in *Advances in Cryptology—CRYPTO 2002 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2002, vol. 2442, pp. 549–564.

[15] ——, "On the secondary constructions of resilient and bent functions," in *Progress in Computer Science and Applied Logic,* K. Feng, H. Niederreiter, and C. Xing, Eds. Cambridge, MA: Birkhäser, 2004, vol. 23, pp. 3–28.

[16] ——, "On bent and highly nonlinear balanced/resilient functions and their algebraic immunities," in *Proc. AAECC 16 (Lecture Notes in Computer Science)*, Las Vegas, NV, 2006, vol. 3857, pp. 1–28.

[17] ——, "Concatenating indicators of flats for designing cryptographic functions," *Des., Codes Cryptogr.*, vol. 36, no. 2, pp. 189–202, 2005.

[18] ——, "On the higher order nonlinearities of algebraic immune functions," in *Proc. CRYPTO 2006 (Lecture Notes in Computer Science)*, to be published.

[19] ——, "Boolean functions for cryptography and error correcting codes," in *Boolean Methods and Models*. Cambridge, U.K.: Cambridge Univ. Press, 2006, (A preliminary version is available at http://www-rocq.inria.fr/codes/Claude.Carlet/pubs.html.), to be published.

[20] C. Carlet and P. Gaborit, "On the construction of balanced Boolean functions with a good algebraic immunity," in *Proc. BFCA (1st Workshop on Boolean Functions: Cryptography and Applications)*, Rouen, France, Mar. 2005, pp. 1–14.

[21] J. H. Cheon and D. H. Lee, "Resistance of S-boxes against algebraic attacks," in *Workshop on Fast Software Encryption (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, , 2004, vol. 3017, , pp. 83–94.

[22] J. Y. Cho and J. Pieprzyk, "Algebraic attacks on SOBER-t32 and SOBER-128," in *Workshop on Fast Software Encryption (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2004, vol. 3017, pp. 49–64.

[23] N. Courtois and J. Pieprzyk, "Cryptanalysis of block ciphers with overdefined systems of equations," in *Advances in Cryptology—ASIACRYPT 2002 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2002, vol. 2501, , pp. 267–287.

[24] N. Courtois and W. Meier, "Algebraic attacks on stream ciphers with linear feedback," in *Advances in Cryptology—EUROCRYPT 2003 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2003, vol. 2656, pp. 345–359.

[25] N. Courtois, "Fast algebraic attacks on stream ciphers with linear feedback," in *Advances in Cryptology—CRYPTO 2003 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2003, vol. 2729, pp. 176–194.

[26] ——, "Cryptanalysis of SFINKS," in *Proc. ICISC 2005*, 2005, vol. 3935, *Lecture Notes in Computer Science* [Online]. Available: http://eprint.iacr.org/

[27] D. K. Dalai, K. C. Gupta, and S. Maitra, "Results on algebraic immunity for cryptographically significant Boolean functions," in *Indocrypt 2004 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2004, vol. 3348, pp. 92–106.

[28] ——, "Cryptographically significant Boolean functions: Construction and analysis in terms of algebraic immunity," in *Workshop on Fast Software Encryption (FSE 2005) (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2005, vol. 3557, pp. 98–111.

[29] D. K. Dalai, S. Maitra, and S. Sarkar, "Basic theory in construction of Boolean functions with maximum possible annihilator immunity," *Des. Codes, Cryptogr.*, vol. 40, no. 1, pp. 41–58, Jul. 2006, Also, avaialable [Online] at http://eprint.iacr.org/.

[30] D. K. Dalai, K. C. Gupta, and S. Maitra, "Notion of algebraic immunity and its evaluation related to fast algebraic attacks," in *Proc. 2nd Int. Workshop on Boolean Functions: Cryptography and Applications (BFCA 2006)*, Rouen, France, Mar. 2006 [Online]. Available: eprint.iacr.org

[31] C. Ding, G. Xiao, and W. Shan*, The Stability Theory of Stream Ciphers (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1991, vol. 561.

[32] E. Filiol and C. Fontaine, "Highly nonlinear balanced Boolean functions with a good correlation-immunity," in *Advances in Cryptology—EUROCRYPT'98 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1998.

[33] M. Hell, A. Maximov, and S. Maitra, "On efficient implementation of search strategy for rotation symmetric Boolean functions," in *Proc. 9th Int. Workshop on Algebraic and Combinatorial Coding Theory (ACCT 2004)*, Kraveno, Bulgaria, Jun. 2004.

[34] T. Johansson and F. Jönsson, "Fast correlation attacks through reconstruction of linear polynomials," in *Advances in Cryptology—CRYPTO 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1880, pp. 300–315.

[35] D. H. Lee, J. Kim, J. Hong, J. W. Han, and D. Moon, "Algebraic attacks on summation generators," in *Workshop on Fast Software Encryption (FSE 2004) (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2004, vol. 3017, pp. 34–48.

[36] M. Lobanov, Tight Bound Between Nonlinearity and Algebraic Immunity [Online]. Available: http://eprint.iacr.org/

[37] A. Maximov, M. Hell, and S. Maitra, "Plateaued rotation symmetric Boolean functions on odd number of variables," in *Proc. 1st Workshop on Boolean Functions: Cryptography and Applications, BFCA 05*, Rouen, France, Mar. 2005 [Online]. Available: eprint.iacr.org

[38] W. Meier, E. Pasalic, and C. Carlet, "Algebraic attacks and decomposition of Boolean functions," in *Advances in Cryptology—EUROCRYPT 2004 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2004, vol. 3027, pp. 474–491.

[39] D. Olejár and M. Stanek, "On cryptographic properties of random Boolean functions," *J. Universal Comput. Sci.*, vol. 4, no. 8, pp. 705–717, 1998.

[40] E. Pasalic, S. Maitra, T. Johansson, and P. Sarkar, "New constructions of resilient and correlation immune Boolean functions achieving upper bounds on nonlinearity," in *Workshop on Coding and Cryptography—WCC 2001 (Electronic Notes in Discrete Mathematics)*. Amsterdam, The Netherlands: Elsevier Science, 2001, vol. 6.

[41] E. Pasalic, "Degree optimized resilient Boolean functions from Maiorana-McFarland class," in *Proc. 9th IMA Conf. Cryptography and Coding*, Cirencester, U.K., Dec. 2003, vol. 2898, *Lecture Notes in Computer Science*, pp. 93–113.

[42] P. Sarkar and S. Maitra, "Construction of nonlinear Boolean functions with important cryptographic properties," in *Advances in Cryptology—EUROCRYPT 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, May 2000, vol. 1807, pp. 485–506.

[43] ——, "Nonlinearity bounds and construction of resilient Boolean functions," in *Advances in Cryptology—Crypto 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1880, pp. 515–532.

[44] J. Seberry, X. M. Zhang, and Y. Zheng, "On constructions and nonlinearity of correlation immune Boolean functions," in *Advances in Cryptology—EUROCRYPT'93 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1994, pp. 181–199.

[45] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.

[46] T. Siegenthaler, "Correlation-immunity of nonlinear combining functions for cryptographic applications," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 5, pp. 776–780, Sep. 1984.

[47] P. Stănică and S. Maitra, "Rotation symmetric Boolean functions—Count and cryptographic properties," in *R. C. Bose Centenary Sympo. Discrete Mathematics and Applications (Electronic Notes in Discrete Mathematics)*. Berlin, Germany: Springer-Verlag, Dec. 2002, vol. 15.

[48] P. Stănică, S. Maitra, and J. Clark, "Results on rotation symmetric bent and correlation immune Boolean functions," in *Fast Software Encryption 2004 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2004, vol. 3017, pp. 161–177.

[49] Y. V. Tarannikov, "On resilient Boolean functions with maximum possible nonlinearity," in *Progress in Cryptology—INDOCRYPT 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1977, pp. 19–30.

[50] D. Wagner, "A generalized birthday problem," in *Advances in Cryptology—CRYPTO 2002 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2002, vol. 2442, pp. 288–303.

[51] F Didier, A new upper bound on the block error probability after decoding over the erasure channel. [Online]. Available: http://www-rocq.inria.fr/codes/Frederic.Didier/\\. A new revised version will appear in *IEEE Trans. Inf. Theory*