Parameterization of Boolean functions by vectorial functions and associated constructions

Claude Carlet,
University of Bergen, Norway and University of Paris 8 (LAGA), France.

E-mail: claude.carlet@gmail.com

Abstract

Despite intensive research on Boolean functions for cryptography for over thirty years, there are very few known general constructions allowing to satisfy all the necessary criteria for ensuring the resistance against all the main known attacks on the stream ciphers using them as nonlinear components. In this paper, we investigate the general construction of Boolean functions f from vectorial functions, in which the support of f equals the image set of an injective vectorial function F, that we call a parameterization of f. Any Boolean function whose Hamming weight is a power of 2, and in particular, every balanced Boolean function, can be obtained this way. We study five illustrations of this general construction. The three first correspond to known classes of functions (Maiorana-McFarland, majority functions and balanced functions in odd numbers of variables with optimal algebraic immunity). The two last correspond to new classes of Boolean functions:
- the sums of indicators of disjoint graphs of $(k, n-k)$-functions,
- functions parameterized by highly nonlinear injective vectorial $(n-1, n)$-functions derived from functions due to Beelen and Leander.
We study the cryptographic parameters (corresponding to the main criteria) of balanced Boolean functions, according to those of their parameterizations: the algebraic degree of f, that we relate to the algebraic degrees of F and of its graph indicator, the nonlinearity of f, that we relate by a bound to the nonlinearity of F, and the algebraic immunity (AI), whose optimality is related to a natural question in linear algebra, and which may be approached (in two ways) by using the graph indicator of F. We show how the algebraic degree and the nonlinearity of the parameterized function can be controlled. We revisit each of the five classes for each criterion. We show that the fourth class is very promising, thanks to a lower bound on the nonlinearity by means of the nonlinearity of the chosen $(k, n-k)$-functions. Its sub-class made of the sums of indicators of affine functions, for which we prove an upper bound on the nonlinearity, seems also interesting. The fifth class includes functions with an optimal algebraic degree, good nonlinearity and good AI.
We leave for future (mostly hard) works the determination of simple ef-

fective sufficient conditions on F ensuring that f has a good AI, the completion of the study of the fourth class, the mathematical study of the AI and fast algebraic immunity of the functions in the fifth class, and the invention of a class of parameterized functions having good parameters and whose output would be fast to compute.

1 Introduction

The main topic of this paper is Boolean functions for a use in stream ciphers, but we shall also incidentally partially address a more purely algebraic problem: given a positive integer r, what are the bases of the vector space $\mathbb{F}_2^r$ such that there exists a family $\mathcal{F}$ in $\mathbb{F}_2^r$ for which the elements of the basis are all the Hadamard (i.e. coordinatewise) products of at most d elements in $\mathcal{F}$, where $\sum_{i=0}^{d}\binom{|\mathcal{F}|}{i} = r$, and with the convention that the empty product (when no element is involved in it) equals the all-1 vector ? The reply to this question is straightforward for $d = 1$, the families $\mathcal{F}$ being all those completing the singleton $\{(1,\ldots,1)\}$ into a basis. The reply is also clear for r a power of 2, say $r = 2^n$, and $d = |\mathcal{F}| = n$, by viewing then $\mathbb{F}_2^r$ as the Reed-Muller code of length 2^n and order n: the classical basis corresponding to the generator matrix G of this code and given by all n-variable monomials in increasing degrees (some order being chosen for monomials of the same degree) is clearly a solution, and this provides all possible bases having the desired property because, given such basis with $(1,\ldots,1)$ as first element and the elements of $\mathcal{F}$ as elements of indices $2,\ldots,n$, the matrix whose rows are all the elements of the basis is necessarily a generator matrix of this Reed-Muller code, up to coordinate permutation, since each column only depends on its coordinates of indices $2,\ldots,n+1$, and this gives the 2^n columns of G in some order. But except in these two cases, the reply to the question is non-trivial. We shall illustrate this for $r = 2^{n-1}$, where n is odd, $|\mathcal{F}| = n$ and $d = \frac{n-1}{2}$, by showing that such bases correspond to those n-variable Boolean functions whose algebraic immunity (an important cryptographic parameter for Boolean functions used in stream ciphers) is optimal.

Let us now present our main subject. Many stream ciphers use Boolean functions as nonlinear components, either as feedback functions in nonlinear feedback shift registers (NFSR), or as filter functions in the filter model (see e.g. [7]) or in the filter permutator [20, 19]. These functions need to satisfy a series of criteria at the best possible levels. The criteria are still unclear in the former case, but are well understood (and quantified by parameters when relevant) in the two latter cases: balance (and when guess and determine attacks are possible, resiliency), a large algebraic degree, a large nonlinearity, a large algebraic immunity and a large fast algebraic immunity (and, when guess and determine attacks are possible, the same requirements on the so-called descendant functions, obtained by fixing some coordinates). The known general classes of Boolean functions whose parameters can be evaluated are few. The main one is the class of Maiorana-McFarland. But this (primary) construction has a drawback: the algebraic immunity and the fast algebraic immunity are

not quite good. There are other general classes for which all the criteria can be evaluated, see [11], but these classes fit with the particular filter permutator model [20, 19], and are not quite suitable for a use in the filter model. There are smaller classes of Boolean functions for the filter model whose criteria can be evaluated, see [10, 7], but we would like to have more functions at our disposal and, moreover, these functions have a rather specific structure (which may represent a threat if new attacks are found), and are not very fast (recall that stream ciphers are supposed to be faster than block ciphers, and the filter function being the only nonlinear part in them, the speed of the cipher is directly linked to the possibility of computing the output of the filter function in a fast way). For these reasons, new general primary and secondary constructions are needed (the former building functions from scratch and the latter building functions from already defined, and so-called initial, functions). The purpose of the present paper is to study the rather natural primary construction consisting in using that the support of any n-variable Boolean function f whose Hamming weight is a power of 2, say equals 2^k (in particular, the support of any balanced Boolean function, in which case $k = n-1$), can be obtained as the image set of an injective function F from $\mathbb{F}_2^k$ to $\mathbb{F}_2^n$. We shall call such F a *parameterization* of f.

We study the main cryptographic criteria of Boolean functions f (algebraic degree, nonlinearity, algebraic immunity) in terms of related properties of their parameterizations F. We show how the algebraic degree of f is related to that of F and of its graph indicator, and that it is easy to ensure that f has an optimal algebraic degree. We prove a bound involving the nonlinearities of f and F, which implies that it is sufficient to take a vectorial $(n, n-1)$-injection with a good nonlinearity for designing a balanced Boolean function with a good nonlinearity, and we observe that, for n odd and $k = n-1$, function f has optimal algebraic immunity if and only if the coordinate functions of F and all the products of at most $\frac{n-1}{2}$ of them form a basis of the vector space $\mathcal{B}_{n-1}$ of all $(n-1)$-variable Boolean functions. This makes a connection with the question asked at the beginning of the present introduction. We also characterize in several ways the non-existence of nonzero annihilators of algebraic degree strictly less than d of a given balanced function in terms of the graph indicators of its parameterizations, and we make clear the situation of functions with optimal algebraic immunity. We study five examples of classes of balanced functions (Maiorana-McFarland functions, majority functions, balanced functions in odd numbers of variables with optimal algebraic immunity, sums of graph indicators of $(k, n-k)$-functions and functions derived from some highly nonlinear vectorial functions due to Beelen and Leander [2]). The first two classes are given as illustrations of the general construction. The third class corresponds to a well-known characterization of Boolean functions with optimal algebraic immunity in odd numbers of variables. The fourth class is very promising, thanks to an interesting lower bound on the nonlinearity by means of the nonlinearities of the $(k, n-k)$-functions; its sub-class where the $(k, n-k)$-functions are affine is also quite interesting (for different reasons, since this lower bound does not give then information). Studying this class will need a whole paper that we

plan for the future. The last class is also quite interesting, since we can control the algebraic degree of the resulting functions, which have also automatically good nonlinearity, and since computer investigations show that a good algebraic immunity can be reached with them.

The paper is organized as follows. After preliminaries, we introduce formally in Section 3 and study the parameterizations of Boolean functions of Hamming weight 2^k; we study the five classes mentioned above. In Section 4, we express the criteria and parameters for such Boolean functions according to their parameterizations, and we continue the study of the five classes. We end with a conclusion in which we draw perspectives.

2 Preliminaries

In this paper, we shall denote the same way, by $+$, additions in $\mathbb{F}_2$, in $\mathbb{F}_2^n$, in $\mathbb{F}_{2^n}$, and in $\mathbb{R}$, since there will be no ambiguity. We shall denote by 0 the zero vector in any of the vector spaces over $\mathbb{F}_2$ and when needing to specify, we shall denote by 0_n the zero vector of length n. We shall also denote by 1_n the all-1 vector of length n. We call *n-variable Boolean function* every function from $\mathbb{F}_2^n$ to $\mathbb{F}_2$ and we denote by $\mathcal{B}_n$ the vector space of all n-variable Boolean functions. The *support* of a Boolean function f is the set $supp(f) = \{x \in \mathbb{F}_2^n;\ f(x) = 1\}$, while the support of a vector $x \in \mathbb{F}_2^n$ equals $\{i \in \{1, \ldots, n\}; x_i = 1\}$. We call *co-support*s the complements of the supports. The *Hamming weight* $w_H(f)$ of a Boolean function f (or of a vector) equals the size of its support. An n-variable Boolean function is called balanced if it has Hamming weight 2^{n-1}. The *Hamming distance* between two Boolean functions f, g is $d_H(f, g) = w_H(f + g)$. The functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ are called *(n, m)-function*s. Such function F being given, the *n-variable* Boolean functions $f_1, \ldots, f_m$, defined at every $x \in \mathbb{F}_2^n$ by $F(x) = (f_1(x), \ldots, f_m(x))$, are called the *coordinate functions* of F. When the numbers m and n are not specified, (n, m)-functions are called *vectorial Boolean functions* or simply *vectorial functions*. Balanced vectorial functions are those such that every pre-image of an element in the co-domain has the same size. Those ones whose role is to ensure confusion in a block cipher are called *substitution boxes* (*S-boxes*). We refer to [7] for a complete state of the art. Two vectorial functions F and G are called affine equivalent if there exist two affine permutations L over $\mathbb{F}_2^m$ and L' over $\mathbb{F}_2^n$ such that $G = L \circ F \circ L'$.

Among the classical representations of Boolean functions and of vectorial functions are the *truth-table* in the case of Boolean functions and the *look-up table* (LUT) in the case of vectorial functions. Both are the table of all pairs of an element of $\mathbb{F}_2^n$ (an ordering of $\mathbb{F}_2^n$ being fixed) and of the value of the function at this input. The *algebraic normal form* (in brief the *ANF*), which contains a little more information directly usable on the cryptographic strengths of functions, is the unique n-variable multivariate polynomial representation of

the form

$$F(x) = \sum_{I \subseteq \{1,\ldots,n\}} a_I \left(\prod_{i \in I} x_i \right) = \sum_{I \subseteq \{1,\ldots,n\}} a_I \, x^I, \tag{1}$$

where a_I belongs to $\mathbb{F}_2$ in the case of Boolean functions and to $\mathbb{F}_2^m$ in the case of (n,m)-functions (and where "x^I" is a notation that we shall use all along this paper). Note that we can deduce the ANF of the i-th coordinate function of F by replacing in (1) each coefficient $a_I \in \mathbb{F}_2^m$ by its i-th coordinate.
The degree of the ANF shall be denoted by $d_{alg}(f)$ (resp. $d_{alg}(F)$); it is called the *algebraic degree* of the function and equals $\max\{|I|;\, a_I \neq 0\}$, where $|I|$ denotes the size of I (with the convention that the zero function has algebraic degree 0). This makes sense thanks to the existence and uniqueness of the ANF. Note that the algebraic degree of an (n,m)-function F equals the maximal algebraic degree of its coordinate functions. It also equals the maximal algebraic degree of its *component functions*, that is, of the nonzero linear combinations over $\mathbb{F}_2$ of the coordinate functions, *i.e.* the functions of the form $v \cdot F$, where $v \in \mathbb{F}_2^m \setminus \{0\}$ and "$\cdot$" is an inner product in $\mathbb{F}_2^m$. It is an affine invariant, that is, its value is preserved by *affine equivalence* (two functions F and G being called affine equivalent if $G = A \circ F \circ A'$ where A and A' are two affine permutations). We have:

$$\forall x \in \mathbb{F}_2^n,\ F(x) = \sum_{I \subseteq supp(x)} a_I, \tag{2}$$

which is valid for Boolean and vectorial functions, and where $supp(x)$ denotes the support of x.
The converse is also true: for all $I \subseteq \{1,\ldots,n\}$, we have:

$$\forall I \subseteq \{1,\ldots,n\},\ a_I = \sum_{x \in \mathbb{F}_2^n;\ supp(x) \subseteq I} F(x), \tag{3}$$

which is also valid for Boolean and vectorial functions. According to Relation (3), we have the well known property (see [18, 7]):

Proposition 1 *An n-variable Boolean function f satisfies $d_{alg}(f) = n$ if and only if $w_H(f)$ is odd. More generally, an (n,m)-function F satisfies $d_{alg}(F) = n$ if and only if $\sum_{x \in \mathbb{F}_2^n} F(x) \neq 0_m$.*

The affine (Boolean or vectorial) functions are the functions of algebraic degree at most 1. We call *quadratic* the Boolean or vectorial functions of algebraic degree at most 2. A vectorial function F is balanced if and only if every of its component functions is balanced.
Recall (see [18]) that for every m and every $r \leq m$, the *Reed-Muller code* $RM(m,r)$ of length 2^m and order r equals the vector space of all m-variable Boolean functions of algebraic degree at most r. It admits as a basis the family of all monomials $x^I = \prod_{i \in I} x_i$ of degree $|I| \leq r$. Its dimension is then $k = \sum_{i=0}^{r} \binom{m}{i}$. Each m-variable Boolean function being associated with a binary vector of length 2^m called its image vector, $RM(m,r)$ is in fact a vector

subspace of $\mathbb{F}_2^{2^m}$, admitting (by definition) as generator matrix any $k \times 2^m$ matrix whose rows constitute a basis of this vector space. For instance we can take as generator matrix of $RM(m,r)$ the matrix whose rows are the image vectors of the monomials above, with the first row equal to the all-1 vector, that is, the image vector of the constant monomial, and if these monomials are written in ascending degree, we have that the rows of this generator matrix equal the Hadamard (i.e. coordinatewise) products of the rows of indices $2, \ldots, n+1$.

The so-called *univariate representation* of an (n,n)-function is in some cases a more convenient representation, obtained after identification between the vector space $\mathbb{F}_2^n$ and the finite field $\mathbb{F}_{2^n}$: the latter being an n-dimensional vector space over $\mathbb{F}_2$, let $(e_1, \ldots, e_n)$ be a basis of this vector space, then any $x \in \mathbb{F}_2^n$ can be viewed as $\sum_{j=1}^n x_j e_j \in \mathbb{F}_{2^n}$, that we shall still denote by x. Then (see e.g. [7]) there is a unique representation of F in the form

$$F(x) = \sum_{i=0}^{2^n-1} a_i x^i \quad \in \quad \mathbb{F}_{2^n}[x]/(x^{2^n}+x)$$

with $a_i \in \mathbb{F}_{2^n}$. For instance, the univariate representation of the Dirac (or Kronecker) function (whose only nonzero value is at 0_n and equals 1) is $\delta_0(x) = x^{2^n-1}+1$. The simplest example of a Boolean function in univariate representation is the so-called *absolute trace function* $tr_n(x) = x+x^2+x^{2^2}+\cdots+x^{2^{n-1}}$. It is valued in $\mathbb{F}_2$, since it satisfies $(tr_n(x))^2 = tr_n(x^2) = tr_n(x)$, and is $\mathbb{F}_2$-linear. Every linear form over $\mathbb{F}_{2^n}$ writes $tr_n(ax)$ where $a \in \mathbb{F}_{2^n}$ is unique.
The *bivariate representation* of n-variable Boolean functions f and of (n,m)-functions F where n is even and $m = \frac{n}{2}$ is as follows: we identify $\mathbb{F}_2^n$ with $\mathbb{F}_{2^m} \times \mathbb{F}_{2^m}$ and we consider then the input to F as an ordered pair (x,y) of elements of $\mathbb{F}_{2^m}$. There exists a unique bivariate polynomial $\sum_{0 \le i,j \le 2^m-1} a_{i,j} x^i y^j$ over $\mathbb{F}_{2^m}$ such that the given function is the bivariate polynomial function over $\mathbb{F}_{2^m}$ associated to it.
We shall call *polynomial representations* all the ways (univariate, bivariate, etc.) of representing vectorial functions in fields of order larger than 2 (there are indeed other possibilities of representing vectorial functions in polynomial representation as we shall see when studying Class 5 below).

The *Fourier-Hadamard transform* of any pseudo-Boolean function φ (i.e. any function from $\mathbb{F}_2^n$ to $\mathbb{R}$) is the $\mathbb{R}$-linear mapping which maps φ to the function $\widehat{\varphi}$ defined on $\mathbb{F}_2^n$ by

$$\widehat{\varphi}(u) = \sum_{x \in \mathbb{F}_2^n} \varphi(x)\,(-1)^{u \cdot x}, \quad u \in \mathbb{F}_2^n, \tag{4}$$

where "$\cdot$" is some chosen inner product in $\mathbb{F}_2^n$. It satisfies the so-called *inverse Fourier-Hadamard transform formula*: for all $a \in \mathbb{F}_2^n$, we have:

$$\sum_{u \in \mathbb{F}_2^n} \widehat{\varphi}(u)\,(-1)^{u \cdot a} = 2^n \varphi(a),$$

which proves that the Fourier-Hadamard transform is a bijection.
If L is an $\mathbb{F}_2$-linear automorphism of $\mathbb{F}_2^n$ and $a \in \mathbb{F}_2^n$, and if L' is the adjoint

operator of L^{-1}, defined by $L'(u) \cdot x = u \cdot L^{-1}(x)$ for every $x, u \in \mathbb{F}_2^n$ (and whose matrix is the transpose of that of L^{-1}in the case the inner product is the so-called usual one $u \cdot x = \sum_{i=1}^n u_i x_i$), the Fourier-Hadamard transform of the function $\varphi'(x) = \varphi(L(x) + a)$, that is, $\widehat{\varphi'}(u) = \sum_{x \in \mathbb{F}_2^n} \varphi'(x) (-1)^{u \cdot x}$ is equal to $\sum_{x \in \mathbb{F}_2^n} \varphi(x) (-1)^{u \cdot L^{-1}(x+a)} = (-1)^{u \cdot L^{-1}(a)} \sum_{x \in \mathbb{F}_2^n} \varphi(x) (-1)^{L'(u) \cdot x} = (-1)^{L'(u) \cdot a} \widehat{\varphi}(L'(u))$.

Given an n-variable Boolean function f (we shall address vectorial functions below), we have two associated transforms: the Fourier-Hadamard transform of f, where f is then viewed as a function from $\mathbb{F}_2^n$ to $\{0, 1\} \subset \mathbb{Z}$, and the *Walsh transform* of f which is the Fourier-Hadamard transform of the sign function $(-1)^f$:

$$W_f(u) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) + u \cdot x}.$$

We have:

$$W_f = 2^n \, \delta_0 - 2\widehat{f}, \tag{5}$$

where δ_0 denotes the Dirac (or Kronecker) symbol over $\mathbb{F}_2^n$, whose ANF is:

$$\delta_0(x) = \prod_{i=1}^{n} (x_i + 1) = \sum_{I \subseteq \{1, \ldots n\}} x^I. \tag{6}$$

The Walsh transform allows to characterize the so-called resilient functions. An n-variable Boolean function f is called t-resilient for some $t \leq n$ if any restriction of f obtained by fixing the coordinates of its input at $n - t$ fixed positions is balanced (see e.g. [7]). This is equivalent to: $W_f(u) = 0$ for every u of Hamming weight at most t.

The *nonlinearity* of a Boolean function f is the minimum Hamming distance between f and affine Boolean functions. We shall denote it by $nl(f)$. We have:

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} |W_f(a)|. \tag{7}$$

The nonlinearity should be large for allowing a resistance of the stream ciphers using f as a filter function to resist fast correlation attacks. The so-called covering radius bound states:

$$nl(f) \leq 2^{n-1} - 2^{\frac{n}{2}-1}. \tag{8}$$

A function is called bent if this inequality is an equality.

Let f be any n-variable Boolean function. An n-variable Boolean function g such that $fg = 0$ is called an *annihilator* of f.
The minimum algebraic degree of nonzero annihilators of f or $f+1$ is called the *algebraic immunity* of f and is denoted by $AI(f)$. It also equals the minimal value d such that there exist $g \neq 0$ and h, both of algebraic degree at most d, such that $fg = h$. The algebraic immunity should be large for allowing a

resistance of the stream ciphers using f as a filter function to resist algebraic attacks. We have $AI(f) \leq \max(d_{alg}(f), \lceil \frac{n}{2} \rceil)$. We say that f has optimal algebraic immunity if $AI(f) = \lceil \frac{n}{2} \rceil$ and almost optimal algebraic immunity if $AI(f) = \lceil \frac{n}{2} \rceil - 1$. For n odd, the functions with optimal algebraic immunity are necessarily balanced.

We address now vectorial functions. We call *Walsh transform* of an (n, m)-*function* F, and we denote by W_F, the function which maps any ordered pair $(u, v) \in \mathbb{F}_2^n \times \mathbb{F}_2^m$ to the value at u of the Walsh transform of the Boolean function $v \cdot F$ (by abuse of notation, we denote similarly the inner products in $\mathbb{F}_2^n$ and $\mathbb{F}_2^m$):

$$W_F(u, v) = \sum_{x \in \mathbb{F}_2^n} (-1)^{v \cdot F(x) + u \cdot x}; \ u \in \mathbb{F}_2^n, v \in \mathbb{F}_2^m.$$

The *nonlinearity* of an (n, m)-function is the minimum nonlinearity of its component functions $v \cdot F$, $v \neq 0$:

$$nl(F) = 2^{n-1} - \frac{1}{2} \max_{\substack{v \in \mathbb{F}_2^m \setminus \{0_m\} \\ u \in \mathbb{F}_2^n}} |W_F(u, v)| . \tag{9}$$

The so-called Sidelnikov-Chabaud-Vaudenay (SCV) bound improves upon the covering radius bound when $m \geq n$. For every (n, m)-function F, it states that $nl(F) \leq 2^{n-1} - \frac{1}{2}\sqrt{3 \times 2^n - 2 - 2\frac{(2^n-1)(2^{n-1}-1)}{2^m-1}}$ (and is achieved with equality if and only if $m = n$ is odd, in which case the bound gives $nl(F) \leq 2^{n-1} - 2^{\frac{n-1}{2}}$). We shall be interested in the case $m = n + 1$, since $nl(F)$ will play a role in the nonlinearity of a related m-variable function (and we shall then be interested in its value by means of m); the bound writes then: $nl(F) \leq 2^{m-2} - \frac{1}{2}\sqrt{3 \times 2^{m-1} - 2 - 2\frac{(2^{m-1}-1)(2^{m-2}-1)}{2^m-1}} = 2^{m-2} - \frac{1}{2}\sqrt{\frac{5 \times 2^{2m-2} - 2^{m+1}}{2^m-1}}$.

Almost bent functions, which exist for every n odd (and can be permutations), are those (n, n)-functions for which the SCV bound is an equality; they satisfy $W_F(u, v) \in \{0, 2^{\frac{n-1}{2}}\}$ for every $u, v \in \mathbb{F}_2^n$, $v \neq 0_n$. They are almost perfect nonlinear. An (n, n)-function F is called almost perfect nonlinear (APN) if, for every nonzero $a \in \mathbb{F}_2^n$ and every $b \in \mathbb{F}_2^n$, the equation $D_aF(x) = F(x) + F(x + a) = b$ has at most two solutions in $\mathbb{F}_2^n$, that is, has either two solutions or none (see e.g. [8, 7]). Function D_aF is called a *derivative* of F. APN functions contribute optimally to the resistance against differential attacks when they are used as S-boxes in block ciphers.

We call *graph indicator* of an (n, m)-function F the $(n+m)$-variable Boolean function equal to the indicator (i.e. the characteristic function) $1_{\mathcal{G}_F}$ of the graph $\mathcal{G}_F = \{(x, F(x)); x \in \mathbb{F}_2^n\}$. We have $1_{\mathcal{G}_F}(x, y) = 1$ if $F(x) = y$ and $1_{\mathcal{G}_F}(x, y) = 0$ otherwise, and then:

$$1_{\mathcal{G}_F}(x, y) = \prod_{j=1}^{m} (y_j + f_j(x) + 1). \tag{10}$$

The properties of $1_{\mathcal{G}_F}$ have been studied in [5, 6].
The Walsh transform of F equals the Fourier-Hadamard transform of $1_{\mathcal{G}_F}$, where

the chosen inner product is $(x,y)\cdot(u,v) = x\cdot u + y\cdot v$. Then, for all $(u,v) \neq (0,0)$, we have:

$$W_F(u,v) = -\frac{1}{2}W_{1_{\mathcal{G}_F}}(u,v),$$

and we have:

$$W_F(0,0) = 2^{2n-1} - \frac{1}{2}W_{1_{\mathcal{G}_F}}(0,0) = 2^n.$$

3 Parameterized Boolean functions

We introduce now the formal definition of the parameterizations of Boolean functions.

Definition 1 *Let F be an injective (k,n)-function and f a Boolean function. We say that f is* parameterized by F *and we denote then f by f_F if $supp(f) = Im(F) := \{F(z), z \in \mathbb{F}_2^k\}$. Then z is called a* parameter *of the Boolean function, k the* parameter dimension *and F a* parameterization *of f_F.*

Then f_F has Hamming weight 2^k. This is why we can write "the" parameter dimension; we write "a parameterization" because every n-variable Boolean function of Hamming weight 2^k has $(2^k)!$ parameterizations. Note that the Boolean functions parameterized by affine functions are the indicators of affine spaces in $\mathbb{F}_2^n$, that is, the minimum weight elements of Reed-Muller codes.

Remark. The interest of considering such parameterizations is twofold: firstly, the approach by parameterization shall lead to new constructions of functions suitable for use in stream ciphers (we shall see examples) and may lead to others; secondly, the parameterization helps in some cases studying the cryptographic parameters of the functions (see Section 4). ◇

More generally, if we do not assume that F is injective, we can define f_F as follows:

Definition 2 *Let F be a (k,n)-function and f a Boolean function. We say that f is* oddly-parameterized by F *if $supp(f) = \{x \in \mathbb{F}_2^n; |F^{-1}(x)|$ is odd$\}$, where $F^{-1}(x) = \{z \in \mathbb{F}_2^{n-1}; F(z) = x\}$.*

This definition allows, as we shall see, to easily determine the ANF and polynomial representations of f from those of F (of course, we could also define evenly-parameterized functions, but they would simply be the complements of oddly-parameterized functions).
In the sequel, when we shall consider a function f_F corresponding to a possibly non-injective vectorial function F, it will always be defined as in Definition 2. The Hamming weight of such function can be different from a power of 2. In fact, any Boolean function of even Hamming weight can be obtained this way. Some results in the present paper will be valid for the general class of oddly

parameterized functions and some will work only for the subclass of parameterized functions.

Remark. Another option would be that F is an (n,n)-function that is 2-to-1 (i.e. is such that every element in $\mathbb{F}_2^n$ has either two pre-images or none) and the support of f equals the image set of F. This option deserves a paper of its own (but we shall give an example of illustration when generalizing Class 4 in Subsection 3.2). The 2-to-1 functions have been studied in [22, 17]. $\diamond$

Although the question of the existence of a k-dimensional parameterization for a given Boolean function f does not pose any problem (since it is equivalent to the fact that $w_H(f) = 2^k$), finding an explicit parameterization for a given class of functions needs work. We observe in the next subsection that this is an affine invariant problem and we study examples in the subsequent subsection.

3.1 Impact of an affine change of parameterization

Clearly, two (k,n)-functions are the parameterizations of a same Boolean function (of Hamming weight 2^k) if and only if one equals the other composed on the right by a permutation of $\mathbb{F}_2^k$ and/or on the left by a permutation of $\mathbb{F}_2^n$ preserving the image set.
Changing F into an affine equivalent function $L \circ F \circ L'$, where L is an affine automorphism of $\mathbb{F}_2^n$ and L' an affine automorphism of $\mathbb{F}_2^k$, transforms f_F into an affine equivalent Boolean function, since the composition by L' does not change f_F (nor does the composition by a nonlinear permutation) and:

$$f_{L\circ F} = f_F \circ L^{-1}.$$

Indeed, given a Boolean function f and a permutation π (affine or not), we have $f \circ \pi^{-1}(x) = 1$ if and only if $x \in \pi(supp(f))$ and the support of $f \circ \pi^{-1}$ equals then the image by π of the support of f.

3.2 On the parameterization of three known classes of balanced Boolean functions and the introduction of two new ones

We begin with a preliminary observation useful for studying Class 1 below. A basic example of a balanced Boolean function has the form $f(x_1,\ldots,x_n) = x_n + g(x_1,\ldots,x_{n-1})$. Such function admits as a parameterization the $(n-1,n)$-function defined as $F(z_1,\ldots,z_{n-1}) = (z_1,\ldots,z_{n-1},g(z_1,\ldots,z_{n-1})+1)$. Indeed, we have $f(x) = 1$ if and only if $x_n = g(x_1,\ldots,x_{n-1}) + 1$, that is, x belongs to the image set of F. Such vectorial function F is clearly injective since its output is the concatenation of that of identity and of that of a Boolean function. Its image set being included in the support of f, it equals this support. We know (see [18, 7]) that any balanced quadratic function has such form, up to affine equivalence. This provides an easy parameterization of any balanced quadratic

function (and in particular of any non-constant affine Boolean function).

Class 1 (Concatenations of parameterized functions; particular case of Maiorana-McFarland's functions). An important way (so-called a secondary construction) of constructing Boolean functions from already built Boolean functions is concatenation. The concatenation $f(x,y) = f_y(x)$; $x \in \mathbb{F}_2^n, y \in \mathbb{F}_2^m$, of parameterized functions f_y having the same parameter dimension k, is a parameterized function of parameter dimension $k+m$: for every $y \in \mathbb{F}_2^m$, let F_y be a parameterization of f_y, then since $supp(f)$ equals $\bigcup_{y\in\mathbb{F}_2^m}(supp(f_y)) \times \{y\} = \bigcup_{y\in\mathbb{F}_2^m}(Im(F_y)) \times \{y\}$, then f admits the parameterization $F(z,y) = (F_y(z),y) \in \mathbb{F}_2^n \times \mathbb{F}_2^m$; $(z,y) \in \mathbb{F}_2^k \times \mathbb{F}_2^m$. Every Maiorana-McFarland's function $f(x,y) = x\cdot\phi(y)+h(y)$ (see [7]), where $\phi(y) \neq 0$ for every $y \in \mathbb{F}_2^m$, enters in this framework since it is a concatenation of non-constant affine functions, (and so does more generally every concatenation of balanced quadratic functions, see [7] as well). We study this class because Maiorana-McFarland are important functions, but we know that their algebraic immunity is somewhat their weak point.$\diamond$

Class 2 (Majority function). Let n be odd (handling the case of n even is similar but slightly more technical). The majority function takes value 1 if and only if its input $x \in \mathbb{F}_2^n$ has Hamming weight at least $\frac{n+1}{2}$. For every $z \in \mathbb{F}_2^{n-1}$ such that $w_H(z) \geq \frac{n-1}{2}$, we can take $F(z)$ equal to $(z,1)$ (the concatenation of z and of the bit 1), and for every $z \in \mathbb{F}_2^{n-1}$ such that $w_H(z) \leq \frac{n-3}{2}$, we can take $F(z)$ equal to $(z+1_{n-1},0)$, where 1_{n-1} is the all-1 vector of length $n-1$. Note that the majority function has optimal algebraic immunity but has low nonlinearity as we shall recall. We take it as an example for illustrating the parameterization principle. $\diamond$

Class 3 (Balanced functions in odd numbers of variables with optimal algebraic immunity). Let us consider the (self-dual) Reed-Muller code $RM(n, \frac{n-1}{2})$ (see e.g. [18, 7]). Since its dimension equals $\sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} = 2^{n-1}$, its generator matrix G is a $2^{n-1} \times 2^n$ matrix of rank 2^{n-1}. We shall assume that its first row is the all-1 vector and the next n rows are the image vectors of the coordinate functions $x_1, \ldots, x_n$. The other rows are the Hadamard (i.e. coordinatewise) products of the rows of indices $2 \ldots, n+1$ (see Section 2). It is well-known as reported in [7], that all Boolean functions in an odd number of variables and with optimal algebraic immunity are balanced, and (see [3], also reported in [7]) that any balanced n-variable Boolean function f has optimal algebraic immunity $\frac{n+1}{2}$ if and only if the matrix obtained by selecting those columns of G whose sub-vectors obtained by keeping the coordinates of indices $2, \ldots, n+1$ belong to the support of f has full rank 2^{n-1}. Let us then choose 2^{n-1} linearly independent columns of G (for instance obtained by Gauss reduction) and call f the balanced n-variable Boolean function whose support equals the set of vectors located in these 2^{n-1} columns at the rows of indices $2, \ldots n+1$. We obtain a $2^{n-1} \times 2^{n-1}$ binary matrix G' whose rows are the Hadamard products of its rows of indices $2 \ldots, n+1$. These latter rows can be

viewed as the lists of values taken by n Boolean functions $f_1, \ldots, f_n$ in $n-1$ variables: indexing the 2^{n-1} columns of the matrix by $z \in \mathbb{F}_2^{n-1}$, the value of $f_i(z)$ equals the entry of the matrix that is located at the $(i+1)th$ row and the column of index z. We have, denoting by F the $(n-1,n)$-function whose coordinate functions are these n functions, that the function f_F equals f (and has optimal algebraic immunity) because each element in the support of f corresponds to a value $z \in \mathbb{F}_2^{n-1}$ (coding the position of the corresponding column in matrix G') and its coordinates are $f_1(z), \ldots, f_n(z)$. We can view this as a method for building all Boolean functions of algebraic immunity $\frac{n+1}{2}$ (but it is more theoretical than practical). $\diamond$

We introduce now new classes, which we shall study in detail in the sequel:

Class 4 (sums of disjoint graph indicators). Let us first consider the case where function f is the graph indicator $1_{\mathcal{G}_G}$ of a single function G. For f to be an n-variable function, we must take for G a $(k, n-k)$-function. Then $1_{\mathcal{G}_G}$ equals f_F where $F(z) = (z, G(z))$. Function f_F has then Hamming weight 2^k and if we want it balanced, we need to take $k = n-1$. Function G is then Boolean and writing g instead of G (as it is usual with Boolean functions), the support of $f_F(z,y)$ has equation $y = g(z)$, with $y \in \mathbb{F}_2$ and $z \in \mathbb{F}_2^{n-1}$, and this gives $f_F(x,y) = y + g(x) + 1$, leading to a simple affine extension of the $(n-1)$-variable Boolean function g into an n-variable function. This secondary construction has little interest.
Considering now the general case of a Boolean function f whose support equals the union of the graphs of $(l, n-l)$-functions G_t, where $t \in \mathbb{F}_2^{n-l-1}$ (so that f is balanced), satisfying $G_t(s) \neq G_{t'}(s)$ for every $s \in \mathbb{F}_2^l$ and every $t \neq t' \in \mathbb{F}_2^{n-l-1}$ (that is, the function $t \mapsto G_t(s)$ is injective for every s), function f is then balanced and equals f_F where, for $z = (s,t)$, we have $F(z) = (s, G_t(s))$. Clearly, if we put no other constraint on the functions G_t, an n-variable function f has this form if and only if, for every $s \in \mathbb{F}_2^l$, the number of $y \in \mathbb{F}_2^{n-l}$ such that the concatenated vector (s,y) belongs to the support of f equals 2^{n-l-1}. This condition looks like lth-order resiliency but it is in fact much lighter: f would be l-resilient if for every $I \subset \{1, \ldots, n\}$ of size l and every $s \in \mathbb{F}_2^I$, there were 2^{n-l-1} elements in the support of f whose coordinates x_i match those of s for $i \in I$. Here we take only $I = \{1, \ldots, l\}$. Note that Class 4 is invariant under complementation, in the sense that if f is a balanced function equal to the sum of disjoint graph indicators, then $f+1$ is one too (which is useful for studying the algebraic immunity). Note also that Class 4 contains all the functions in Class 1 which are concatenations of balanced functions, that is, $f(s,x) = f_s(x)$; $s \in \mathbb{F}_2^k, x \in \mathbb{F}_2^{n-k}$, where f_s is balanced, and in particular, all the Maiorana-McFarland functions $f(s,x) = x \cdot \phi(s) + h(s)$ where ϕ does not vanish.
The simplest option when choosing functions G_t is to take them affine. It is not clear which constraint this puts on f (the resulting condition should not be confused with the fact that the 2^{n-l-1} elements in the support of f whose coordinates x_i match those of s for $i \in \{1, \ldots, l\}$ range in an affine space). The constraint on each G_t is, given a basis $(e_1, \ldots, e_l)$ of $\mathbb{F}_2^l$, that, for every

$\epsilon = (\epsilon_1, \ldots, \epsilon_l) \in \mathbb{F}_2^l$, we have $G_t(\sum_{i=1}^l \epsilon_i e_i) = G_t(0) + \sum_{i=1}^l \epsilon_i (G_t(e_i) + G_t(0))$, but it is hard to say what is the resulting condition on f, since given s, it is difficult to know, for each y such that $(s, y) \in supp(f)$, what is the value of t such that $y = G_t(s)$. This leaves some freedom for the constructions of such functions f. We consider it an advantage rather than a drawback. Now, if we are given some balanced Boolean function f, find such affine G_t may be challenging (unless l is small). But building f after choosing the functions G_t is easy and this is what needs to be done for obtaining a construction.
This class and this latter particular case would deserve a whole paper for being studied in details. It can be made more general by considering Boolean functions whose supports equal the disjoint union of permuted graphs of $(l, n-l)$-functions G_t, that is, denoting the symmetric group over $\{1, \ldots, n\}$ by $\mathcal{S}_n$: $supp(f_F) = \bigcup_{t \in \mathbb{F}_2^{n-l-1}} \{\pi_t(s, G_t(s)), s \in \mathbb{F}_2^l\}$, where $\pi_t \in \mathcal{S}_n$. But this generalization may be complex to study because of the condition that the permuted graphs are disjoint.
Another generalization may be more interesting, in which we do not impose anymore that the function $t \mapsto G_t(s)$ is injective for every s but that it is for instance 2-to-1 (and the support of f is still the union of the disjoint image sets of these 2-to-1 mappings). Then t would no more live in $\mathbb{F}_2^{n-l-1}$ but in $\mathbb{F}_2^{n-l}$. This is (almost) the case for instance when f is a so-called γ_F function related to an almost perfect nonlinear function, as defined in [8]. We have already recalled that an (m, m)-function F is called almost perfect nonlinear if, for every nonzero $a \in \mathbb{F}_2^m$, there are at most two solutions $x \in \mathbb{F}_2^m$ to the equation $F(x) + F(x + a) = b$. Then we have $\gamma_F(a, b) = 1$ if $a \neq 0$ and the equation does have solutions. The support of γ_F equals then the union of the graphs of the 2-to-1 functions $x \mapsto D_a F(x) = F(x) + F(x + a)$ for a nonzero in $\mathbb{F}_2^m$ (a playing the role of s and x playing the role of t and $G_t(s)$ being equal to $D_a F(x)$)). Function γ_F is not balanced but if we replace the zero function $b \mapsto \gamma_F(0, b)$ by a balanced Boolean function, say g, we obtain a function which enters in the framework of the present generalization, with $m = l = n - l = \frac{n}{2}$. $\diamond$

Class 5 (Functions having a modified Beelen-Leander vectorial function for parameterization). We shall see in Subsection 4.3 that the nonlinearity of F plays a role in the nonlinearity of f_F, and that starting with a highly nonlinear $(n-1, n)$-function F is a good way of obtaining a Boolean function f_F with good nonlinearity. A highly nonlinear $(n-1, n)$-function has been introduced, for every n even, by Beelen and Leander in [2]. Denoting by H the linear hyperplane of $\mathbb{F}_{2^{\frac{n}{2}}}$ equal to $\{t \in \mathbb{F}_{2^{\frac{n}{2}}}; tr_{\frac{n}{2}}(at) = 0\}$, where $tr_{\frac{n}{2}}$ is the trace function introduced in Section 2 and a is such that $tr_{\frac{n}{2}}(a) = 1$, this function is defined over $\mathbb{F}_{2^{\frac{n}{2}}} \times H$ (which has dimension $\frac{n}{2} + \frac{n}{2} - 1 = n - 1$, as needed) as follows:

$$(z, t) \mapsto \left(\frac{z^2}{t+1}, \frac{z^3}{(t+1)^2} \right); \tag{11}$$
$$z \in \mathbb{F}_{2^{\frac{n}{2}}}, t \in H = \{t \in \mathbb{F}_{2^{\frac{n}{2}}}; tr_{\frac{n}{2}}(at) = 0\}.$$

Unfortunately, this function is not injective, since every input $(0, t)$ maps to

$(0,0)$. However, restricted to $\mathbb{F}^*_{2^{\frac{n}{2}}} \times H$, the function is injective since we can recover z and t from $(x,y) = \left(\frac{z^2}{t+1}, \frac{z^3}{(t+1)^2}\right)$ by the relations $z = \frac{x^2}{y}$ and $t+1 = \frac{x^3}{y^2}$. Then, adding for instance $\delta_0(z)(G(t),0)$ to (x,y), where δ_0 is the Dirac (Kronecker) symbol and G is an injective function from H to $\mathbb{F}_{2^{\frac{n}{2}}}$, changes the Beelen-Leander function into an injective function F. We could also try to add a well-chosen linear $(n-1,n)$-function so as to obtain an injection, since this would not change the nonlinearity, but it seems more difficult to find such a suitable linear function (and most vectorial functions cannot be changed into injective ones by the addition of affine functions). ◇

Remark The fact that n is even is a limitation in the generality (however, practically, n even is preferred in cryptography, being more convenient for computability). The limits of a paper do not allow us to study more examples which would allow n to be odd. We would need an infinite class of highly nonlinear $(n-1,n)$-functions, which would need to be modifiable into injective functions. In [2], Beelen and Leander construct $(2r-1,lr)$-functions (we denote here by l what they denote by $k-2$) and this provides $(n-1,n)$-functions only for $l=2$ and $n=2r$. Now, we could also take $l>2$ and discard coordinate functions of the highly nonlinear vectorial function, since this does not reduce the nonlinearity. The lower bound on the nonlinearity that we could derive would be worse since their functions have worse nonlinearity for larger l, but it would not be much worse (indeed the lower bound obtained by Beelen and Leander is $2^{n-1} - 2^{r-2}(l+1)$, which is linear in l) and maybe the actual values could be good.
Another way of obtaining highly nonlinear injective mappings is to take highly nonlinear $(n-1,n-1)$-permutations and add one well chosen coordinate function. But such $(n-1,n)$-functions F lead to functions f_F whose support has the form $\{(z,h(z)); z \in \mathbb{F}_2^{n-1}\}$, and which satisfy then $f_F(x) = x_n + h(x') + 1$, where $x' = (x_1,\ldots,x_{n-1})$ and are then simple affine extensions of $(n-1)$-variable Boolean functions. ◇

4 The representations and main cryptographic properties of parametrized functions

4.1 Algebraic normal form and algebraic degree

Let F be an injective (k,n)-function, given by its ANF. For every $x \in \mathbb{F}_2^n$, there exists at most one $z \in \mathbb{F}_2^k$ such that $x = F(z)$, that is, $\delta_0(x+F(z)) = 1$. We have then, according to Relation (6):

$$f_F(x) = \sum_{z \in \mathbb{F}_2^k} \delta_0(x+F(z)) = \sum_{z \in \mathbb{F}_2^k} \Big(\prod_{i=1}^{n} (x_i + f_i(z) + 1) \Big), \tag{12}$$

where f_i is the i-th coordinate function of F. These expressions are also valid for oddly parameterized functions. The ANF of f_F can then be obtained by

expanding and simplifying Relation (12) and using Proposition 1:

$$f_F(x) = \sum_{\substack{I\subseteq\{1,\ldots,n\}\\ z\in\mathbb{F}_2^k}} \left(\prod_{i\in I^c}(f_i(z)+1)\right) x^I = \sum_{\substack{I\subseteq\{1,\ldots,n\}\\ d_{alg}(\prod_{i\in I^c}(f_i(z)+1))=k}} x^I, \tag{13}$$

where $I^c = \{1,\ldots,n\}\setminus I$.
We deduce the value of the algebraic degree of f_F:

$$d_{alg}(f_F) = \max\Big\{|I|; I\subseteq\{1,\ldots,n\}, d_{alg}\Big(\prod_{i\in I^c}(f_i(z)+1)\Big) = k\Big\}. \tag{14}$$

Remark. Since $\prod_{i\in I^c}(f_i(z)+1)$ equals $\sum_{J\subseteq I^c}\prod_{i\in J} f_i(z)$, we can replace the integer $d_{alg}(\prod_{i\in I^c}(f_i(z)+1))$ by $d_{alg}(\prod_{i\in I^c} f_i(z))$ in this latter relation.

Proposition 2 *Let F be any (k,n)-function. The algebraic degree of the oddly parameterized function f_F equals $n-l$ where l is the minimum number of coordinate functions of F whose product has algebraic degree k (that is, odd Hamming weight). In particular, we have $d_{alg}(f_F) = n-1$ if and only if $d_{alg}(F) = k$, that is, $\sum_{z\in\mathbb{F}_2^k} F(z) \neq 0_n$.*

Hence, we have a way of ensuring that f_F has optimal algebraic degree: we just need to take F of optimal algebraic degree. For addressing the general case, let us now provide an expression of the value of $d_{alg}(f_F)$, which is alternative to (14), and leads to an upper bound which will show that the graph indicator $1_{\mathcal{G}_F}$ of F (that is, the Boolean function whose support equals the graph $\mathcal{G}_F = \{(z,F(z)); z\in\mathbb{F}_2^k\}$ of F) must have large enough algebraic degree for allowing f_F to have large algebraic degree.

Proposition 3 *Let F be any (k,n)-function and let $1_{\mathcal{G}_F}$ be the graph indicator of F. The algebraic degree of the oddly parameterized function f_F equals:*

$$d_{alg}(f_F) = n - \min\{d_{alg}(g); d_{alg}(g\circ F) = k\}, \tag{15}$$

and satisfies:

$$d_{alg}(f_F) \leq d_{alg}(1_{\mathcal{G}_F}) - k. \tag{16}$$

Proof. We know (see e.g. [7]) that, as for any Boolean function, the algebraic degree of f_F equals $n-d$, where d is the minimum algebraic degree of those n-variable Boolean functions g such that $\sum_{x\in\mathbb{F}_2^n} g(x)f_F(x) \neq 0$. In the present case, this results in $\sum_{z\in\mathbb{F}_2^k} g(F(z)) = 1$, that is, $d_{alg}(g\circ F) = k$. This proves (15). It is shown in [6] (and reported in [7, Relation (2.9)]) that $d_{alg}(g\circ F) \leq d_{alg}(1_{\mathcal{G}_F}) + d_{alg}(g) - n$. The condition $d_{alg}(g\circ F) = k$ implies then $d_{alg}(g) \geq k + n - d_{alg}(1_{\mathcal{G}_F})$. This completes the proof. □

We know that $d_{alg}(1_{\mathcal{G}_F})$ is located strictly between n and $n+k$, see [5, 6]). We shall have with Class 2 an example where Inequality (16) is an equality.

Class 1 (continued). There is no point of studying the ANF of Class 1 nor its algebraic degree, since these functions were defined by their ANF. Let us study the ANF of their parameterization functions. In the case of function $x_n + g(x_1, \ldots, x_{n-1})$ we have seen that we can take $F(z_1, \ldots, z_{n-1}) = (z_1, \ldots, z_{n-1}, g(z_1, \ldots, z_{n-1}) + 1)$, which provides the ANF of F. We have seen in Subsection 3.1 that composing on the right a Boolean function by an affine automorphism corresponds to composing its parameterization by the inverse of this affine automorphism, on the left. According to what we have seen on concatenation at Subsection 3.2, concatenating functions of the form $x_n + g_y(x_1, \ldots, x_{n-1})$ composed on the right by an automorphism L_y, gives then the parameterization:

$$F : (z_1, \ldots, z_{n-1}, y) \mapsto (L_y^{-1}(z_1, \ldots, z_{n-1}, g_y(z_1, \ldots, z_{n-1}) + 1), y).$$

This formula is valid for Maiorana-McFarland functions. The algebraic degrees of F and $1_{\mathcal{G}_F}$ depend on the correspondence between y and L_y. $\diamond$

Class 2 (continued). The ANF of the majority function is known (it is recalled for instance in [7, Subsection 10.1.7]): for n odd, the coefficient of x^I equals $\binom{|I|-1}{\frac{n-1}{2}}$ (mod 2). The algebraic degree is determined in [13] and equals $2^{\lfloor \log_2(n) \rfloor}$. Let us determine the ANF of the parameterization F which has been presented at Subsection 3.2. For every $i = 1, \ldots, n-1$, the i-th coordinate function $f_i(z)$ of $F(z)$ equals z_i if $w_H(z) \geq \frac{n-1}{2}$ and $z_i + 1$ otherwise; the last coordinate function $f_n(z)$ equals 1 if $w_H(z) \geq \frac{n-1}{2}$ and 0 otherwise. Let $t_{n-1,\frac{n-1}{2}}(z)$ be the threshold function in $n-1$ variables whose value at z equals 1 if and only if $w_H(z) \geq \frac{n-1}{2}$ (in other words, $t_{n-1,\frac{n-1}{2}}(z)$ is the majority function in $n-1$ variables). The coefficient of z^I in the ANF of $t_{n-1,\frac{n-1}{2}}(z)$ equals $\binom{|I|-1}{\frac{n-3}{2}}$ (mod 2). For every $i = 1, \ldots, n-1$, we have then $f_i(z) = z_i\, t_{n-1,\frac{n-1}{2}}(z) + (z_i + 1)(t_{n-1,\frac{n-1}{2}}(z) + 1) = z_i + t_{n-1,\frac{n-1}{2}}(z) + 1$ and we have $f_n(z) = t_{n-1,\frac{n-1}{2}}(z)$. The algebraic degree of F equals then that of $t_{n-1,\frac{n-1}{2}}$, that is, $2^{\lfloor \log_2(n-1) \rfloor}$. Finally, we have $1_{\mathcal{G}_F}(z, x) = \prod_{j=1}^{n}(x_j + f_j(z) + 1) = (x_n + t_{n-1,\frac{n-1}{2}}(z) + 1) \prod_{j=1}^{n-1}(x_j + z_j + t_{n-1,\frac{n-1}{2}}(z))$ and, n being odd, the algebraic degree of $1_{\mathcal{G}_F}$ equals then $2^{\lfloor \log_2(n) \rfloor} + n - 1$. We can see that the bound of Proposition 3 is then an equality. $\diamond$

Class 3 (continued). Recall from Subsection 3.2 that, given the classical generator matrix G of the Reed-Muller code $RM(n, \frac{n-1}{2})$, and a non-singular $2^{n-1} \times 2^{n-1}$ sub-matrix G' of G, the rows of indices 2 to $n+1$ of G' are viewed as n Boolean functions $f_1, \ldots, f_n$ in $n-1$ variables and provide the coordinate functions of a parameterization F of the n-variable balanced function f_F with optimal algebraic immunity in odd number of variables whose support is the set of indices (which are elements of $\mathbb{F}_2^n$) of the columns of G chosen for G'. Assume that we are given this support $S \subset \mathbb{F}_2^n$, then the ANF of f is obtained by applying Relation (3), which writes $\forall I \subseteq \{1, \ldots, n\}$, $a_I = |\{x \in S;\ supp(x) \subseteq I\}|$ (mod 2), and the ANF of a possible parameterization F is

obtained as follows: denoting by $u^{(1)}, \ldots, u^{(2^{n-1})}$ the elements of S, we have that $F(z)$ equals $u^{(\bar{z}+1)}$ where $\bar{z}$ is the integer whose binary expansion equals z, that is, $F(z) = u^{(1+\sum_{j=1}^{n-1} z_j 2^{j-1})}$. Indeed, this function is injective and its image set equals the support of f. But the expression $F(z) = u^{(1+\sum_{j=1}^{n-1} z_j 2^{j-1})}$ is not the ANF of F since z intervenes in an exponent. The ANF of F writes $F(z) = \sum_{t \in \mathbb{F}_2^{n-1}} u^{(1+\sum_{j=1}^{n-1} t_j 2^{j-1})} \prod_{j=1}^{n-1}(z_j + t_j + 1)$. The ANF of the graph indicator of F is then given by Relation (10). The algebraic degrees of F and of its graph indicator depend then in a rather complex way on the $u^{(i)}$.

Class 4 (continued). The support of f in this class being the union of the graph indicators of $(l, n-l)$-functions G_t, where $t \in \mathbb{F}_2^{n-l-1}$ and $G_t(s) \neq G_{t'}(s)$ for every $s \in \mathbb{F}_2^l$ and every $t \neq t' \in \mathbb{F}_2^{n-l-1}$, we have, denoting by $g_1^t, \ldots, g_{n-l}^t$ the coordinate functions of G_t and using Relation (10), that function $f = f_F$ where $F(s,t) = (s, G_t(s))$, satisfies:

$$f(x,y) = \sum_{t \in \mathbb{F}_2^{n-l-1}} \prod_{i=1}^{n-l}(y_i + g_i^t(x) + 1);\ x \in \mathbb{F}_2^l, y \in \mathbb{F}_2^{n-l}.$$

Using Proposition 2 (last sentence of) and that $\sum_{s \in \mathbb{F}_2^l} s = 0$, we have that $d_{alg}(f) = n-1$ if and only if $\sum_{s \in \mathbb{F}_2^l, t \in \mathbb{F}_2^{n-l-1}} G_t(s) \neq 0$.
According to Relation (10), the ANF of $1_{\mathcal{G}_F}$ writes

$$1_{\mathcal{G}_F}(s,t,x,y) = \prod_{j=1}^{l}(x_j + s_j + 1) \prod_{i=1}^{n-l}(y_i + g_i^t(s) + 1),$$

with $(s,t,x,y) \in \mathbb{F}_2^l \times \mathbb{F}_2^{n-l-1} \times \mathbb{F}_2^l \times \mathbb{F}_2^{n-l}$. We have then $d_{alg}(1_{\mathcal{G}_F}) \leq l + d_{alg}\Big(\prod_{i=1}^{n-l}(y_i + g_i^t(s) + 1)\Big)$. Assuming that for each i, the function $g_i^t : (s,t) \mapsto g_i^t(s)$ has algebraic degree at least 1, we have then $d_{alg}(1_{\mathcal{G}_F}) \leq l + \sum_{i=1}^{n-l} d_{alg}(g_i^t)$ and Proposition 3 (with $k = n-1$) implies that $d_{alg}(f) \leq 1 + \sum_{i=1}^{n-l}(d_{alg}(g_i^t) - 1)$. $\diamond$

4.2 Polynomial representation

There is no natural univariate representation of a (k,n)-function when $k < n$, unless k divides n (but this does not interest us much since f_F cannot then be balanced).

We have however seen with Class 5 that other polynomial representations of (k,n)-functions exist. Let us then assume that we have a representation of a (k,n)-function F whose coefficients and variables live in a field $\mathbb{F}_{2^m}$, which can be a sub-field of $\mathbb{F}_{2^n}$ as in the case of the Beelen-Leander function (where $m = \frac{n}{2}$). Then a polynomial representation of f_F can be derived from the

relation:

$$f_F(x) = \sum_{z \in \mathbb{F}_2^k} \delta_0(x + F(z)), \tag{17}$$

where δ_0 is the Dirac symbol in $\mathbb{F}_2^n$. The polynomial representation of δ_0 can be different according to what is m. Over $\mathbb{F}_{2^n}$, we have $\delta_0(x) = x^{2^n-1} + 1$ as we already saw; over $\mathbb{F}_{2^{\frac{n}{2}}}$, we have $\delta_0(x) = x^{2^{\frac{n}{2}}-1} + 1$; and over $\mathbb{F}^2_{2^{\frac{n}{2}}}$, we have $\delta_0(x, y) = (x^{2^{\frac{n}{2}}-1} + 1)(y^{2^{\frac{n}{2}}-1} + 1)$.

Class 5 (continued). Recall that we have $F(z,t) = \left(\frac{z^2}{t+1} + \delta_0(z)\, G(t), \frac{z^3}{(t+1)^2}\right) = \left(\frac{z^2}{t+1} + (z^{2^{\frac{n}{2}}-1} + 1)\, G(t), \frac{z^3}{(t+1)^2}\right)$, where $z \in \mathbb{F}_{2^{\frac{n}{2}}}, t \in H = \{t \in \mathbb{F}_{2^{\frac{n}{2}}};\, tr_{\frac{n}{2}}(at) = 0\}$ and where G is an injective function from H to $\mathbb{F}_{2^{\frac{n}{2}}}$. Then, function $f_F(x, y)$ equals:

$$\sum_{\substack{z \in \mathbb{F}_{2^{n/2}} \\ t \in H}} \left(\left(x + \frac{z^2}{t+1} + (z^{2^{\frac{n}{2}}-1} + 1)\, G(t)\right)^{2^{\frac{n}{2}}-1} + 1\right)\left(\left(y + \frac{z^3}{(t+1)^2}\right)^{2^{\frac{n}{2}}-1} + 1\right).$$

It seems difficult to deduce the algebraic degree from this polynomial representation. But we can use Proposition 2. For $n \geq 6$, the two functions $z \mapsto z^2$ and $z \mapsto z^3$ having algebraic degree strictly less than $\frac{n}{2}$, they sum to zero over $\mathbb{F}_{2^{\frac{n}{2}}}$ and we have then $\sum_{z \in \mathbb{F}_{2^{\frac{n}{2}}}, t \in H} F(z,t) = \left(\sum_{t \in H} G(t), 0_{\frac{n}{2}}\right)$. If function G is taken such that $\sum_{t \in H} G(t)$ is nonzero (which is possible with an injective $(\frac{n}{2} - 1, \frac{n}{2})$-function), that is, with algebraic degree $\frac{n}{2} - 1$ over H, then $\sum_{z \in \mathbb{F}_{2^{\frac{n}{2}}}, t \in H} F(z,t)$ is nonzero and f_F has algebraic degree $n - 1$.

4.3 Walsh transform, balance, nonlinearity and resiliency

4.3.1 Walsh transform, balance and nonlinearity

Let F be an injective (k, n)-function. We have $\widehat{f_F}(u) = \sum_{x \in supp(f_F)} (-1)^{u \cdot x} = \sum_{z \in \mathbb{F}_2^k} (-1)^{u \cdot F(z)} = W_F(0_k, u)$, for every $u \in \mathbb{F}_2^n$, and then, according to (5):

$$W_{f_F}(u) = 2^n \delta_0(u) - 2\, W_F(0_k, u). \tag{18}$$

Hence:

Proposition 4 *Let F be an injective (k, n)-function. Then :*

- *If $k \leq n - 2$, then: $nl(f_F) = w_H(f_F) = 2^k$.*
- *If $k = n - 1$, then:*

$$nl(f_F) = 2^{n-1} - \max_{u \in \mathbb{F}_2^n; u \neq 0_n} |W_F(0_{n-1}, u)| \tag{19}$$

$$\geq 2\, nl(F). \tag{20}$$

- *If $k = n$, then $f_F = 1$ and $nl(f_F) = 0$.*

Proof. If $k \leq n-2$, then the Hamming distance $w_H(f_F)$ between f_F and the zero function is smaller than or equal to its Hamming distance to any nonzero affine function h (i.e. f_F is a coset leader of the first-order Reed-Muller code), since by the triangular inequality, we have that if h is non-zero, then $d_H(f_F, h) \geq w_H(h) - w_H(f_F) \geq 2^{n-1} - w_H(f_F) \geq 2^{n-2} \geq w_H(f_F)$.
If $k = n-1$, then

$$\begin{aligned} nl(f_F) &= 2^{n-1} - \max_{u \in \mathbb{F}_2^n} |W_{f_F}(u)| \\ &= 2^{n-1} - \max_{u \in \mathbb{F}_2^n; u \neq 0_n} |W_F(0_{n-1}, u)| \\ &\geq 2^{n-1} - \max_{\substack{u \in \mathbb{F}_2^n; u \neq 0_n \\ t \in \mathbb{F}_2^{n-1}}} |W_F(t, u)| = 2\, nl(F). \end{aligned}$$

The assertion for $k = n$ is straightforward. □

The interesting case, in which we will place ourselves in the rest of this subsection, is of course $k = n-1$. Choosing F with a nonlinearity near the Sidelnikov-Chabaud-Vaudenay (SCV) bound $2^{n-2} - \frac{1}{2}\sqrt{\frac{5 \times 2^{2n-2} - 2^{n+1}}{2^n - 1}}$ (which cannot be reached, but may be approached) ensures according to Proposition 4 that f_F has a nonlinearity near $2^{n-1} - \sqrt{5}\ 2^{\frac{n}{2}-1}$ or larger, which is rather good (and revives the interest in finding $(n-1, n)$-functions approaching the SCV bound). We can even hope reaching in the future with parameterized functions a very good nonlinearity, that is, a nonlinearity near $2^{n-1} - 2^{\frac{n}{2}-1} - 2^{\frac{n}{4}-1} - \cdots - 2^{n'-1} - 2^{\frac{n'-1}{2}}$, where n' is the largest odd divisor of n. This latter value is reached by balanced functions designed by Dobbertin in [15] (and was conjectured optimal by him), but we know that none of the functions constructed in this paper can satisfy all the criteria needed for the function to be used in stream ciphers (see more in [7]) and it would be a huge step forward to reach such good nonlinearity while ensuring good algebraic degree and algebraic immunity.

Remark. Conversely, it would be interesting to study those injective $(n-1, n)$-functions F such that the highly nonlinear balanced functions built by Dobbertin in [15] equal f_F. Some may be highly nonlinear. ◇

Remark. We have seen that, given a Boolean function of Hamming weight 2^k for some k, there are $(2^k)!$ possible choices of the parameterization F. These different choices can have different nonlinearities. In Proposition 4, in the case $k = n-1$, the inequality of Relation (20) can be replaced by

$$nl(f_F) \geq 2 \max_{\pi \in \mathcal{S}_{2^{n-1}}} nl(F \circ \pi), \tag{21}$$

where $\mathcal{S}_{2^{n-1}}$ is the symmetric group over $\mathbb{F}_2^{n-1}$. Note that Relation (21) also

writes:

$$nl(f_F) \geq 2^{n-1} - \max\Big(\max_{u\in\mathbb{F}_2^n;u\neq 0_n} |W_F(0_n,u)|, \min_{\pi\in\mathcal{S}_{2^{n-1}}} \max_{\substack{u\in\mathbb{F}_2^n;u\neq 0_n\\ t\in\mathbb{F}_2^{n-1};t\neq 0_{n-1}}} |W_{F\circ\pi}(t,u)|\Big),$$

and it is an equality if and only if $\min_{\pi\in\mathcal{S}_{2^{n-1}}} \max_{\substack{u\in\mathbb{F}_2^n;u\neq 0_n\\ t\in\mathbb{F}_2^{n-1};t\neq 0_{n-1}}} |W_{F\circ\pi}(t,u)| \leq \max_{u\in\mathbb{F}_2^n;u\neq 0_n} |W_F(0_n,u)|$. This seems to be often the case (but determining whether it is always the case or not seems difficult), thanks to this minimum taken over all permutations π. For instance, if F is linear, then denoting by F^* the adjoint operator of F (satisfying $u\cdot F(z) = F^*(u)\cdot z$), we have:

$$\begin{aligned}
&\min_{\pi\in\mathcal{S}_{2^{n-1}}} \max_{\substack{u\in\mathbb{F}_2^n;u\neq 0_n\\ t\in\mathbb{F}_2^{n-1};t\neq 0_{n-1}}} |W_{F\circ\pi}(t,u)| = \\
&\min_{\pi\in\mathcal{S}_{2^{n-1}}} \max_{\substack{u\in\mathbb{F}_2^n;u\neq 0_n\\ t\in\mathbb{F}_2^{n-1};t\neq 0_{n-1}}} \Big|\sum_{z\in\mathbb{F}_2^{n-1}} (-1)^{F^*(u)\cdot z + t\cdot\pi^{-1}(z)}\Big| = \\
&\min_{\pi\in\mathcal{S}_{2^{n-1}}} \max_{\substack{u\in\mathbb{F}_2^n;u\neq 0_n\\ t\in\mathbb{F}_2^{n-1};t\neq 0_{n-1}}} |W_{\pi^{-1}}(F^*(u),t)| = \\
&\min_{\pi\in\mathcal{S}_{2^{n-1}}} \max_{\substack{w\in\mathbb{F}_2^{n-1};w\neq 0_n\\ t\in\mathbb{F}_2^{n-1};t\neq 0_{n-1}}} |W_{\pi^{-1}}(w,t)|,
\end{aligned}$$

since, F being injective, the functions $z\mapsto u\cdot F(z) = F^*(u)\cdot z$ cover all linear forms over $\mathbb{F}_2^{n-1}$ when u ranges over $\mathbb{F}_2^n$), and F^* is then onto $\mathbb{F}_2^{n-1}$. This minimum is (much) smaller than 2^{n-1}, since, for every n, there are permutations π such that this latter maximum is not much larger than $2^{\frac{n}{2}}$ (since for $n-1$ odd, almost bent permutations have a maximum still smaller, and for $n-1$ even, we know that there are permutations such that this maximum is not far from $2^{\frac{n}{2}}$), while $\max_{u\in\mathbb{F}_2^n;u\neq 0_n} |W_F(0_n,u)|$ equals 2^{n-1}, because the image set of F is an affine hyperplane. Then, (21) writes $nl(f_F)\geq 0$, while $nl(f_F)$ is indeed 0 since f_F is affine. We leave open the interesting but seemingly difficult study of $\min_{\pi\in\mathcal{S}_{2^{n-1}}} \max_{\substack{u\in\mathbb{F}_2^n;u\neq 0_n\\ t\in\mathbb{F}_2^{n-1};t\neq 0_{n-1}}} |W_{F\circ\pi}(t,u)|$ for general injective $(n-1,n)$-functions F and its comparison with $\max_{u\in\mathbb{F}_2^n;u\neq 0_n} |W_F(0_n,u)|$. ◇

Remark. Little work exists in the literature on the nonlinearity of (k,n)-functions when $n>k$. Relation (20) shows that it is interesting to find injective $(n-1,n)$-functions whose nonlinearity is as close to the SCV bound as possible. We have used for defining Class 5 a function from a class (introduced in [2]) of $(2e-1,(j-2)e)$-functions F with nonlinearity $2^{e-2}(2^e-j+1)$ based on Reed-Solomon codes, where $e\geq 2$ and $3\leq j\leq 2^e$. For $j=4$ and $n=2e$, this provided the $(n-1,n)$-function (11) of nonlinearity $nl(F) = 2^{n-2} - 3\cdot 2^{\frac{n}{2}-2}$, that we used for Class 5. More classes of highly nonlinear $(n-1,n)$-functions would be interesting to find. ◇

Classes 1 and 2 (continued). The nonlinearity of Maiorana-McFarland's functions has been much studied, see [7]. That of the majority function is

known, see [13, 7], and equal to $2^{n-1} - \binom{n-1}{\lceil \frac{n-1}{2} \rceil}$, which is weak. We shall then not study Classes 1 and 2 in this subsection.

Class 3 (continued). The nonlinearity of function $f = f_F$ defined in Class 3 equals $\min_{h \in RM(1,n)} d_H(f,h) = \min_{h \in RM(1,n)} w_H(f+h)$, which, since f is balanced, equals $2 \min_{h \in RM(1,n) \setminus RM(0,n)} w_H(f(h+1))$, where $RM(0,n)$ and $RM(1,n)$ are the 0-th order and the first-order Reed-Muller codes of length 2^n, that is, the vector spaces of, respectively, the constant n-variable Boolean functions and the n-variable affine functions; since $RM(1,n)$ is stable under addition of constant function 1, we have then $nl(f) = 2 \min_{h \in RM(1,n)) \setminus RM(0,n)} w_H(fh)$. The nonlinearity equals then twice the minimum distance of the first-order Reed-Muller code, punctured at all positions outside the support of f. A few results are known on such minimum distance (see [1, 14, 21]), but in practice, the nonlinearity must be separately studied for each function f.

Class 4 (continued). For $F(s,t) = (s, G_t(s))$, where $t \in \mathbb{F}_2^{n-l-1}$ and where the $(l, n-l)$-functions G_t satisfy $G_t(s) \neq G_{t'}(s)$ for every $s \in \mathbb{F}_2^l$ and every $t \neq t' \in \mathbb{F}_2^{n-l-1}$, we have, for every $u \in \mathbb{F}_2^l$ and $v \in \mathbb{F}_2^{n-l}$ such that $(u,v) \neq (0_l, 0_{n-l})$:

$$W_F((0_l, 0_{n-l-1}), (u,v)) = \sum_{s \in \mathbb{F}_2^l, t \in \mathbb{F}_2^{n-l-1}} (-1)^{u \cdot s + v \cdot G_t(s)} = \sum_{t \in \mathbb{F}_2^{n-l-1}} W_{G_t}(u,v),$$

and according to (19):

$$\begin{aligned}
nl(f_F) &= 2^{n-1} - \max_{(u,v) \in \mathbb{F}_2^l \times \mathbb{F}_2^{n-l} \setminus \{(0_l, 0_{n-l})\}} \Big| \sum_{t \in \mathbb{F}_2^{n-l-1}} W_{G_t}(u,v) \Big| \\
&\geq 2^{n-1} - \max_{(u,v) \in \mathbb{F}_2^l \times \mathbb{F}_2^{n-l} \setminus \{(0_l, 0_{n-l})\}} \sum_{t \in \mathbb{F}_2^{n-l-1}} \Big| W_{G_t}(u,v) \Big| \qquad (22) \\
&\geq 2^{n-1} - \sum_{t \in \mathbb{F}_2^{n-l-1}} \max_{(u,v) \in \mathbb{F}_2^l \times \mathbb{F}_2^{n-l} \setminus \{(0_l, 0_{n-l})\}} \Big| W_{G_t}(u,v) \Big| \qquad (23) \\
&= 2^{n-1} - \sum_{t \in \mathbb{F}_2^{n-l-1}} \big(2^l - 2\, nl(G_t)\big) = 2 \sum_{t \in \mathbb{F}_2^{n-l-1}} nl(G_t).
\end{aligned}$$

Proposition 5 *Let $(G_t)_{t \in \mathbb{F}_2^{n-l-1}}$ be a family of $(l, n-l)$-functions such that, for every $s \in \mathbb{F}_2^l$ and every $t \neq t' \in \mathbb{F}_2^{n-l-1}$, $G_t(s) \neq G_{t'}(s)$. Let f be the resulting n-variable Boolean function in Class 4, as described in Subsection 3.2. We have:*

$$nl(f_F) \geq 2 \sum_{t \in \mathbb{F}_2^{n-l-1}} nl(G_t).$$

We have then a way to reach a large nonlinearity with f_F, by choosing each G_t with a large nonlinearity. Note that both inequalities in (22) and (23) may be far from equalities and this should allow to reach very good nonlinearities.

We have evoked, in Subsection 3.2, the sub-class where all functions G_t are affine. Of course, the inequality $nl(f_F) \geq 2\sum_{t\in\mathbb{F}_2^{n-l-1}} nl(G_t)$ gives then no information. Let us then study this case apart, taking $G_t(s) = L_t(s) + a_t$, where L_t is a linear $(l, n-l)$-function and a_t is an element of $\mathbb{F}_2^{n-l}$ such that $\forall t \neq t', a_t + a_{t'} \notin Im(L_t + L_{t'})$, so that the condition "$G_t(s) \neq G_{t'}(s)$ for every $s \in \mathbb{F}_2^l$ and every $t \neq t' \in \mathbb{F}_2^{n-l-1}$" is satisfied. Then, denoting again by L_t^* the adjoint operator of L_t, we have $W_{G_t}(u,v) = \sum_{s\in\mathbb{F}_2^l}(-1)^{u\cdot s+v\cdot a_t+L_t^*(v)\cdot s} = \begin{cases} 2^l(-1)^{v\cdot a_t} \text{ if } u = L_t^*(v) \\ 0 \text{ otherwise.} \end{cases}$
Therefore, according to Relation (19):

$$nl(f_F) = 2^{n-1} - 2^l \max_{(u,v)\in\mathbb{F}_2^l\times\mathbb{F}_2^{n-l}\setminus\{(0_l,0_{n-l})\}} \Big| \sum_{t\in\mathbb{F}_2^{n-l-1}; u=L_t^*(v)} (-1)^{v\cdot a_t} \Big|.$$

We note a similarity between this formula and the formula giving the nonlinearity of Maiorana-McFarland functions, whose known properties, developed in [4] (and recalled in [7]), can be adapted to this new situation. We have (since $0_l = L_t^*(0_{n-l})$, for every t):

$$\sum_{(u,v)\in\mathbb{F}_2^l\times\mathbb{F}_2^{n-l}\setminus\{(0_l,0_{n-l})\}} \Big(\sum_{t\in\mathbb{F}_2^{n-l-1}; u=L_t^*(v)} (-1)^{v\cdot a_t} \Big)^2 = \tag{24}$$

$$\sum_{(u,v)\in\mathbb{F}_2^l\times\mathbb{F}_2^{n-l}\}} \Big(\sum_{t\in\mathbb{F}_2^{n-l-1}; u=L_t^*(v)} (-1)^{v\cdot a_t} \Big)^2 - 2^{2n-2l-2} =$$

$$\sum_{(u,v)\in\mathbb{F}_2^l\times\mathbb{F}_2^{n-l}} \Big(\sum_{(t,t')\in(\mathbb{F}_2^{n-l-1})^2; u=L_t^*(v)=L_{t'}^*(v)} (-1)^{v\cdot(a_t+a_{t'})} \Big) - 2^{2n-2l-2} =$$

$$\sum_{(t,t')\in(\mathbb{F}_2^{n-l-1})^2} \Big(\sum_{v\in\ker(L_t^*+L_{t'}^*)} (-1)^{v\cdot(a_t+a_{t'})} \Big) - 2^{2n-2l-2} =$$

$$\sum_{\substack{(t,t')\in(\mathbb{F}_2^{n-l-1})^2 \\ a_t+a_{t'}\in(\ker(L_t^*+L_{t'}^*))^\perp}} \Big|\ker(L_t^* + L_{t'}^*)\Big| - 2^{2n-2l-2}, \tag{25}$$

where $(\ker(L_t^* + L_{t'}^*))^\perp = \{a \in \mathbb{F}_2^{n-l}; \forall v \in \ker(L_t^* + L_{t'}^*), v \cdot a = 0\}$. Note that "$u \neq 0_l$ and $v = 0_{n-l}$" implies $\sum_{t\in\mathbb{F}_2^{n-l-1}; u=L_t^*(v)}(-1)^{v\cdot a_t} = 0$ and the sum (24) over $(u,v) \in \mathbb{F}_2^l \times \mathbb{F}_2^{n-l} \setminus \{(0_l, 0_{n-l})\}$ is in fact over $(u,v) \in \mathbb{F}_2^l \times (\mathbb{F}_2^{n-l} \setminus \{0_{n-l}\})$. Moreover, the sum of $(-1)^{v\cdot a_t}$ over the set $\{v \in \mathbb{F}_2^{n-l}, u = L_t^*(v)\}$ is equal to zero if this set is empty or if a_t is not orthogonal to the direction $\ker(L_t^*)$ of this affine space. Using (25) and the fact that the maximum of a sequence of positive numbers is larger than or equal to its arithmetic mean and smaller than or equal to its sum, we deduce the double inequality:

$$\sqrt{\frac{1}{2^n - 2^l}\Big(\sum_{\substack{(t,t')\in(\mathbb{F}_2^{n-l-1})^2 \\ a_t+a_{t'}\in(\ker(L_t^*+L_{t'}^*))^\perp}} \Big|\ker(L_t^* + L_{t'}^*)\Big| - 2^{2n-2l-2}\Big)} \leq$$

$$\max_{(u,v)\in\mathbb{F}_2^l\times\mathbb{F}_2^{n-l}\setminus\{(0_l,0_{n-l})\}}\left|\sum_{t\in\mathbb{F}_2^{n-l-1};u=L_t^*(v)}(-1)^{v\cdot a_t}\right|\leq$$

$$\sqrt{\sum_{\substack{(t,t')\in(\mathbb{F}_2^{n-l-1})^2\\ a_t+a_{t'}\in(\ker(L_t^*+L_{t'}^*))^\perp}}\left|\ker(L_t^*+L_{t'}^*)\right|-2^{2n-2l-2}}.$$

Therefore, according to Proposition 4:

Proposition 6 *Let $(G_t=L_t+a_t)_{t\in\mathbb{F}_2^{n-l-1}}$ be a family of affine $(l,n-l)$-functions such that $\forall t\neq t', a_t+a_{t'}\notin Im(L_t+L_{t'})$. Let f be the resulting n-variable Boolean function in Class 4, as described in Subsection 3.2. We have the following double inequality:*

$$2^{n-1}-2^l\sqrt{\sum_{\substack{(t,t')\in(\mathbb{F}_2^{n-l-1})^2\\ a_t+a_{t'}\in(\ker(L_t^*+L_{t'}^*))^\perp}}\left|\ker(L_t^*+L_{t'}^*)\right|-2^{2n-2l-2}}\leq nl(f)\leq$$

$$2^{n-1}-2^l\sqrt{\frac{1}{2^n-2^l}\Big(\sum_{\substack{(t,t')\in(\mathbb{F}_2^{n-l-1})^2\\ a_t+a_{t'}\in(\ker(L_t^*+L_{t'}^*))^\perp}}\left|\ker(L_t^*+L_{t'}^*)\right|-2^{2n-2l-2}\Big)}.$$

The lower bound is probably far from the actual value (since we are neglecting all but one of the terms in the sum (24)), but the upper bound may be more precise and it leads to a way, which will be developed in another paper, of choosing the L_t's and the a_t's in order to increase the proportion of highly nonlinear functions in the corresponding corpus. An interesting particular case is when the functions $t\mapsto a_t$ and $t\mapsto L_t$ both are linear. ◇

Class 5 (continued). The functions in Class 5 are highly nonlinear by construction, thanks to the bound (20). We have recalled that function (11) has nonlinearity $2^{n-2}-3\cdot 2^{\frac{n}{2}-2}$. Since it is not injective, we added $\delta_0(z)(G(t),0)$ to its image, where δ_0 is the Dirac (Kronecker) symbol and G is an injective function from H to $\mathbb{F}_{2^{\frac{n}{2}}}$. This changed it into an injective function. Adding this expression may modify the Walsh transform values by at most $2|H|=2^{\frac{n}{2}}$, above or below, and then does not decrease its nonlinearity by more than $2^{\frac{n}{2}-1}$ (and may increase it as well). The nonlinearity of f_F is then at least $2^{n-1}-3\cdot 2^{\frac{n}{2}-2}-2^{\frac{n}{2}-1}=2^{n-1}-5\cdot 2^{\frac{n}{2}-2}$ and may be better. In fact, a very good nonlinearity can be reached with these functions.
A few examples of functions f_F where F is constructed as described above are investigated in [12]. Such investigation needs that a choice of G be made since the number $2^{\frac{n}{2}}(2^{\frac{n}{2}}-1)(2^{\frac{n}{2}}-2)(2^{\frac{n}{2}}-3)\cdots(2^{\frac{n}{2}}-2^{\frac{n}{2}-1}+1)$ of injective functions from H to $\mathbb{F}_{2^{\frac{n}{2}}}$ is much too large. Obviously, any injective function from H to $\mathbb{F}_{2^{\frac{n}{2}}}$ is the restriction to H of an $(\frac{n}{2},\frac{n}{2})$-permutation. We have then investigated those functions F obtained when G is the restriction to H of a power permutation over $\mathbb{F}_{2^{n/2}}$, that is, $G(y)=y^d$, where $\gcd(d,2^{\frac{n}{2}}-1)=1$. These functions

provided good nonlinearity but they did not give good algebraic immunity. We then tried $G(y) = y^d + y^{-1}$. Those d which provided injective functions G over H (and therefore injective parameterizations) were not very numerous but they gave good results (precisely, a slightly smaller nonlinearity than with y^d but a much better algebraic immunity); we shall give their AI in the next subsection and we give here their nonlinearity (compared with the covering radius bound $2^{n-1} - 2^{\frac{n}{2}-1}$):

n	$nl(f_F)$	$2^{n-1} - 5 \cdot 2^{\frac{n}{2}-2}$	covering radius bound
8	110	108	120
10	478	472	496
12	1978	1968	2016
14	8056	8032	8128
16	32498	32448	32640

These results are rather good.
We also used for G the restriction to H of a Dickson permutation polynomial (see e.g. [23]), specifically the one of index $2^{n/2}-2$, which is co-prime with 2^n-1 (condition for the Dickson polynomial of index k to be a permutation polynomial over $\mathbb{F}_{2^{n/2}}$) if and only if $\gcd(n/2-1, n) = 1$, that is, n is multiple of 4. Recall that Dickson polynomials can be easily deduced from the recurrence relation $D_i(x) = xD_{i-1}(x) + D_{i-2}(x)$ and the initial values $D_1(x) = x$, $D_2(x) = x^2$. We obtained:

n	$nl(f_F)$	$2^{n-1} - 5 \cdot 2^{\frac{n}{2}-2}$	covering radius bound
8	110	108	120
12	1978	1968	2016
16	32498	32448	32640

These results are then the same as those above. □

4.3.2 Resiliency

Proposition 7 *Let F be an injective $(n-1, n)$-function. Then f_F is t-resilient if and only if the sum (mod 2) of at most t coordinate functions of F is balanced.*

Since we know that f_F is balanced, this is straightforward, according to Relation (18).

4.4 Algebraic immunity

An n-variable function h is an annihilator of f_F if and only if $h(F(z)) = 0$, for every $z \in \mathbb{F}_2^k$, that is, $h \circ F$ is identically zero. It seems then easier to deal with the annihilators of f than with those of $f + 1$, when a parameterization of a Boolean function f is known. Practically, it is desirable to know a parameterization for each of the functions f and $f + 1$.

There is an exception to this: when n is odd, we know from [3] that if a Boolean function is balanced, then it has an optimal algebraic immunity $\frac{n+1}{2}$ if and only if it admits no nonzero annihilator of algebraic degree at most $\frac{n-1}{2}$. This is what allowed the introduction of class 3. Unfortunately, this result applies only for functions whose algebraic immunity is optimal in an odd number of variables, while for implementation reasons, n even is better, and even for n odd, we know that the algebraic immunity of some interesting functions such as the hidden weight bit function [24] is not optimal.

Given the ANF of F, determining whether every nonzero Boolean function h of algebraic degree strictly smaller than some positive integer d is such that $h \circ F \neq 0$ (i.e. there exists z such that $F(z)$ belongs to the support of h) results, by considering the ANF of the Boolean function $h \circ F$, in the fact that, denoting by $f_1, \ldots, f_n$ the coordinate functions of F, any nonzero linear combination of the products $\prod_{i \in I} f_i$ for $I \subset \{1, \ldots, n\}$, $|I| < d$, is not equal to the zero function. Note that for n odd and $d = \frac{n+1}{2}$, the number $\sum_{j=0}^{d-1} \binom{n}{j}$ of these products equals the dimension 2^{n-1} of $\mathcal{B}_{n-1}$. Hence:

Proposition 8 *For every n, let F be any injective $(n-1, n)$-function and let f_F be the Boolean function parameterized by F.*
1. There is no nonzero annihilator of f_F of algebraic degree strictly less than d if and only if the family of all the products of strictly less than d coordinate functions of F is $\mathbb{F}_2$-linearly independent, with the convention that the empty product (when no function is involved in it) equals constant function 1.
2. For n odd, function f_F has (optimal) algebraic immunity $\frac{n+1}{2}$ if and only if the family of all the products of at most $\frac{n-1}{2}$ coordinate functions of F is a basis of $\mathcal{B}_{n-1}$, with the same convention.

The introduction of Class 3 in Subsection 3.2 is the constructive version of Item 2.

Remark. For n odd, it is then equivalent to find an n-variable Boolean function of optimal algebraic immunity and to find a basis of $\mathcal{B}_{n-1}$, that is, of $\mathbb{F}_2^{2^{n-1}}$, whose elements are obtained as the Hadamard products of at most $\frac{n-1}{2}$ vectors chosen in a family of n binary vectors of length 2^{n-1}. Such basis has a nice structure and is "compact" is the sense that it is enough to store the n vectors for having the whole basis of 2^{n-1} elements. Moreover, for every n-variable Boolean function h, there is according to Item 2 in Proposition 8, a unique choice of $(a_I)_{\substack{I \subseteq \{1,\ldots,n\} \\ |I| \leq \frac{n-1}{2}}} \in \mathbb{F}_2^{2^{n-1}}$, such that $h \circ F$ equals $\sum_{I \subseteq \{1,\ldots,n\}; |I| \leq \frac{n-1}{2}} a_I \prod_{i \in I} f_i$. We can then consider the mapping $h \in \mathcal{B}_n \mapsto \sum_{I \subseteq \{1,\ldots,n\}; |I| \leq \frac{n-1}{2}} a_I x^I$. The kernel of this linear mapping is the space of annihilators of f_F; hence, this linear mapping is a projection over the Reed-Muller code $RM(\frac{n-1}{2}, n)$ of length 2^n and order $\frac{n-1}{2}$, parallel to the space of annihilators of f_F, that is, to the principal ideal of $\mathcal{B}_n$ generated by $f_F + 1$. $\diamond$

4.4.1 An approach with graph indicators

Graph indicators are an important tool for the study of vectorial Boolean functions, see [5], in particular when evaluating and bounding the algebraic degree of composite functions, see [6], as seen in Proposition 3 above. We shall show that they also play a role with the algebraic immunity of parameterized functions. Let us first observe that Proposition 8 can be translated nicely in terms of the graph indicator of F. We have (as we already saw) that:

$$1_{\mathcal{G}_F}(z,x) = \prod_{i=1}^{n}(x_i + f_i(z) + 1) = \sum_{I\subseteq\{1,\ldots,n\}} \prod_{i\in I^c}(f_i(z)+1)x^I, \qquad (26)$$

where $f_1,\ldots,f_n$ are the coordinate functions of F and $I^c = \{1,\ldots,n\}\setminus I$. Note that the non-existence of nonzero annihilators of f_F of degree less than d is preserved when we add a constant to F, since this corresponds to a translation on the support of f_F; hence we can replace each coordinate function f_i by f_i+1. Proposition 8 writes then:

Corollary 1 *Let f be any balanced n-variable Boolean function and F any parameterization of f. Let $1_{\mathcal{G}_F}(z,x) = \sum_{I\subseteq\{1,\ldots,n\}} a_I(z)\,x^I$ be the ANF of the graph indicator of F. Then:*
1. There is no nonzero annihilator of f of algebraic degree strictly less than d if and only if the family $(a_I(z))_{\substack{I\subseteq\{1,\ldots,n\}\\ |I|>n-d}}$ is $\mathbb{F}_2$-linearly independent.
2. For n odd, f has (optimal) algebraic immunity $\frac{n+1}{2}$ if and only if the family $(a_I(z))_{\substack{I\subseteq\{1,\ldots,n\}\\ |I|\geq\frac{n+1}{2}}}$ is a basis of $\mathcal{B}_{n-1}$.

This gives, at least in theory, a way of designing parameterized functions with optimal algebraic immunity:

Corollary 2 *For any odd n, finding all the n-variable Boolean functions of optimal algebraic immunity is equivalent to finding all the Boolean functions $g(z,x) = \sum_{I\subseteq\{1,\ldots,n\}} a_I(z)\,x^I$; $z\in\mathbb{F}_2^{n-1}, x\in\mathbb{F}_2^n$, where $a_I\in\mathcal{B}_{n-1}$ for every $I\subseteq\{1,\ldots,n\}$, such that:*
1. the family $(a_I(z))_{I\subseteq\{1,\ldots,n\};|I|\geq\frac{n+1}{2}}$ is a basis of $\mathcal{B}_{n-1}$,
2. g is a graph indicator, that is, for every $z\in\mathbb{F}_2^{n-1}$, there exists exactly one $x\in\mathbb{F}_2^n$ such that $g(z,x)=1$,
3. the vectorial function whose g is the graph indicator is injective, that is, for every $x\in\mathbb{F}_2^n$, there exists at most one $z\in\mathbb{F}_2^{n-1}$ such that $g(z,x)=1$.

There are two more ways of handling the graph indicator of F.

(i). Keeping the same represention of $1_{\mathcal{G}_F}(z,x)$, using that for every z, there is a unique x such that $1_{\mathcal{G}_F}(z,x)=1$ and this x equals then $F(z)$, and using Proposition 1, we have:

$$h\circ F(z) = \sum_{x\in\mathbb{F}_2^n} 1_{\mathcal{G}_F}(z,x)\,h(x) = \sum_{\substack{I\subseteq\{1,\ldots,n\},\\ x\in\mathbb{F}_2^n}} a_I(z)\,x^I h(x) = \sum_{\substack{I\subseteq\{1,\ldots,n\};\\ d_{alg}(x^I h(x))=n}} a_I(z).$$

Then f_F has no nonzero annihilator of algebraic degree less than d if and only if, for any nonzero function h of algebraic degree less than d, the function $\sum_{\substack{I\subseteq\{1,\ldots,n\}\\ d_{alg}(x^I h(x))=n}} a_I$ is not identically 0. This gives in fact a condition equivalent to that of Corollary 1.

(ii). We can also represent $1_{\mathcal{G}_F}(z,x)$ as follows:

$$1_{\mathcal{G}_F}(z,x) = \sum_{J\subseteq\{1,\ldots,k\}} b_J(x)\, z^J. \tag{27}$$

Then, using the same facts as above in case 1, we have:

$$\begin{aligned} h\circ F(z) = \sum_{x\in\mathbb{F}_2^n} 1_{\mathcal{G}_F}(z,x)\, h(x) &= \sum_{J\subseteq\{1,\ldots,k\}} \left(\sum_{x\in\mathbb{F}_2^n} b_J(x)\, h(x)\right) z^J, \\ &= \sum_{\substack{J\subseteq\{1,\ldots,k\};\\ d_{alg}(b_J\, h)=n}} z^J, \end{aligned}$$

and according to the uniqueness of the representation by the ANF, $h\circ F$ is identically zero if and only if, for every $J\subseteq\{1,\ldots,k\}$, we have $d_{alg}(b_J\, h)<n$, or equivalently, $b_J\, h$ has even Hamming weight, that is, b_J and h are orthogonal relatively to the usual inner product $(h,k)\mapsto \sum_{x\in\mathbb{F}_2^n} h(x)\, k(x)\in\mathbb{F}_2$ in the vector space of n-variable Boolean functions. We have that f_F has no nonzero annihilator of algebraic degree strictly less than d if and only if, for every nonzero Boolean function of algebraic degree strictly less than d (that is, any nonzero element of $RM(d-1,n)$, the Reed-Muller code of order $d-1$ and length 2^n), there exists $J\subseteq\{1,\ldots,k\}$ such that $d_{alg}(b_J\, h)=n$.

Proposition 9 *For every n, let F be any injective (k,n)-function and f_F the indicator function of $Im(F)$.*
The annihilators of f_F are the elements of the orthogonal[1] $<b_J; J\subseteq\{1,\ldots,k\}>^\perp$ *of the $\mathbb{F}_2$-vector space generated by the Boolean functions b_J defined by Relation (27).*
Function f_F has no nonzero annihilator of algebraic degree strictly less than d if and only if all the nonzero elements in $<b_J; J\subseteq\{1,\ldots,k\}>^\perp$ have algebraic degree at least d.
For n odd, function f_F has (optimal) algebraic immunity $\frac{n+1}{2}$ if and only if all the nonzero elements in $<b_J; J\subseteq\{1,\ldots,k\}>^\perp$ have algebraic degree at least $\frac{n+1}{2}$.

Remark. There is a nice similarity with the notion of dual distance of a linear code in coding theory: we have to calculate a dual and to determine its minimum "distance" where, here, the "distance" between two Boolean functions is

[1] In terms of coding theory, $<b_J; J\subseteq\{1,\ldots,k\}>^\perp$ is the linear code whose parity check matrix is made of the lists of values of the functions b_J, written as rows.

the algebraic degree of their difference (that is, their sum). ◇

Remark. This characterization is more or less as complex to handle as that of Corollary 1 (since calculating an orthogonal and showing that all its nonzero elements have algebraic degree at least d is comparable to showing the linear independence of a family of Boolean functions). And determining the polynomials $b_J(x)$ represents more work than determining the $a_I(z)$. Note however that since F is injective, there exists an injective function G such that the function $\bar{F} : (z, z_n) \mapsto \begin{cases} F(z) \text{ if } z_n = 0 \\ G(z) \text{ if } z_n = 1 \end{cases}$ is a permutation. Then, denoting by $\bar{f}'_1(x), \ldots, \bar{f}'_n(x)$ the coordinates of function $\bar{F}^{-1}$, we have for $z_n = 0$ that $1_{\mathcal{G}_F}(z, x) = 1_{\mathcal{G}_{\bar{F}^{-1}}}(x, (z, 0)) = \prod_{j=1}^{n-1}(\bar{f}'_j(x) + z_j + 1)(\bar{f}'_n(x) + 1)$ and we have then:

$$b_J(x) = (\bar{f}'_n(x) + 1) \prod_{j \in \{1,\ldots,n-1\} \setminus J} (\bar{f}'_j(x) + 1).$$

This gives a clue for future research: instead of choosing F, choose a permutation $\bar{F}$ whose inverse is known, and take for F the restriction of $\bar{F}$ to $\mathbb{F}_2^{n-1} \times \{0\}$. Two examples of such permutations obviously come first:
- power permutations: let $(\alpha_1, \ldots, \alpha_n)$ be a basis of the n-dimensional vector space $\mathbb{F}_{2^n}$ and set $\bar{F}(z_1, \ldots, z_{n-1}, z_n) = (\sum_{i=1}^n z_i \alpha_i)^d$ where $\gcd(d, n) = 1$ and where this latter element of $\mathbb{F}_{2^n}$ is identified with the binary vector of its coordinates relatively to the basis $(\alpha_1, \ldots, \alpha_n)$ or to another basis; we have $F(z_1, \ldots, z_n) = (\sum_{i=1}^{n-1} z_i \alpha_i)^d$ where this latter element of $\mathbb{F}_{2^n}$ is identified with the binary vector of its coordinates relatively to the basis $(\alpha_1, \ldots, \alpha_n)$, and $\bar{F}^{-1}(x) = x^{\frac{1}{d}}$, where $\frac{1}{d}$ is the inverse of d in $\mathbb{Z}/(2^n - 1)\mathbb{Z}$; denoting by $(\beta_1, \ldots, \beta_n)$ a dual basis of $(\alpha_1, \ldots, \alpha_n)$, that is a basis such that $tr_n(\alpha_i \beta_j)$ equals 1 if $i = j$ and 0 otherwise, where $tr_n(x) = x + x^2 + x^{2^2} + \cdots + x^{2^{n-1}}$ is the absolute trace function over $\mathbb{F}_{2^n}$, the coordinate functions of $\bar{F}^{-1}$ are $\bar{f}'_j(x) = tr_n(\beta_j x^{\frac{1}{d}})$;
- Dickson polynomials. ◇

4.4.2 The AI of the five classes

Classes 1, 2 and 3 do not need to be studied from the viewpoint of the algebraic immunity since it is well-known that Maiorana-McFarland functions do not have a very good algebraic immunity in general (see e.g. [7]), while the functions in Classes 2 and 3 have an optimal one.

Class 4 (continued). For $F(s, t) = (s, G_t(s))$ (where $G_t(s) \neq G_{t'}(s)$ for every $s \in \mathbb{F}_2^l$ and every $t \neq t' \in \mathbb{F}_2^{n-l-1}$), a Boolean function $h(x, y)$ is an annihilator of f_F if and only if it is an annihilator of the graph indicator of each function G_t, that is, for every t and every $s \in \mathbb{F}_2^l$, $h(s, G_t(s)) = 0$. Assuming that h is nonzero and has algebraic degree at most d, we have $h(x, y) = \sum_{I \subseteq \{1,\ldots,l\}} a_I(x)\, y^I = \sum_{I \subseteq supp(y)} a_I(x)$, where for every $I \subseteq \{1, \ldots, l\}$, a_I is an l-variable Boolean function of algebraic degree at most $d - |I|$, and at least one such a_I is not the zero function. Then h is an annihilator of the

graph indicator of G_t if and only if the $(n-1)$-variable Boolean function $h(s, G_t(s)) = \sum_{I\subseteq\{1,\ldots,l\}} a_I(s)\,(G_t(s))^I = \sum_{I\subseteq supp(G_t(s))} a_I(s)$ equals 0 for every t. The condition for f_F to have no nonzero annihilator of algebraic degree at most d is then that, for any choice of an l-variable Boolean function a_I of algebraic degree at most $d-|I|$, for every $I \subseteq \{1,\ldots,l\}$, one of which is nonzero, there exist $s \in \mathbb{F}_2^l$ and $t \in \mathbb{F}_2^{n-l-1}$ such that the support $S_{t,s}$ of $G_t(s)$ satisfies $\sum_{I\subseteq S_{t,s}} a_I(s) = 1$. This approach will be developed in another paper. It allows to address the annihilators of the sums of disjoint graph indicators, and it covers then also their complements since Class 4 is invariant under complementation.

Class 5 (continued). The methods described above, after being adapted to the particular polynomial representation of the Beelen-Leander function, seem difficult to apply and ad-hoc methods will need to be developed. Indeed, for $z, x, y \in \mathbb{F}_{2^{\frac{n}{2}}}$ and $t \in H$, $1_{\mathcal{G}_F}\Big((z,t),(x,y)\Big)$ equals 1 if and only if $x = \frac{z^2}{t+1} + (z^{2^{n/2}-1}+1)\,G(t)$ and $y = \frac{z^3}{(t+1)^2}$. Hence, we have that $1_{\mathcal{G}_F}\Big((z,t),(x,y)\Big)$ equals:

$$\left(\left(x + \frac{z^2}{t+1} + (z^{2^{n/2}-1}+1)\,G(t)\right)^{2^{n/2}-1} + 1\right)\left(\left(y + \frac{z^3}{(t+1)^2}\right)^{2^{n/2}-1} + 1\right).$$

Applying Corollary 1 or using option (ii) seems hard.
We have computed in [12] the algebraic immunity of the functions in Class 5 whose nonlinearity has been reported in Subsection 4.3. As we already explained, the algebraic immunity when G is a power function being not very good, we tried functions G of the form $G(y) = y^d + y^{-1}$. We obtained the following results:

n	$AI(f_F)$
8	4
10	5
12	5
14	6
16	6

All functions $y^d + y^{-1}$ were not covered for all possible values of a and it is plausible that better results can be found in the future with $y^d + y^{-1}$ and still better AI can be reached with other functions than $y^d + y^{-1}$ (which was, in a way, an arbitrary choice). But a mathematical study may be needed for determining a proper corpus to be investigated. The results already obtained are rather good (with optimal AI for $n = 8, 10$ and almost optimal AI for $n = 12, 14$) but seem to weaken slightly as n increases.
As already mentioned, we also used for G the restriction to H of a Dickson permutation polynomial, specifically the one of index $2^{n/2} - 2$, which is a permutation polynomial when n is multiple of 4. We obtained:

n	$AI(f_F)$
8	4
12	5
16	6

Hence, with algebraic immunity as well, the results are the same. ◇

Conclusion.
We have introduced and studied a natural way of building n-variable balanced Boolean functions from injective $(n-1,n)$-functions (that we call parameterizations of the Boolean functions), in which the support of the Boolean function equals the image set of the vectorial function. An interest of our construction is that, if the parameterization is taken with a large nonlinearity, then the corresponding Boolean function has rather large nonlinearity as well. We have started the study of a derived class of Boolean functions equal to sums of disjoint graph indicators, which needs to be further studied. We have also studied a class whose parameterization is derived from highly nonlinear vectorial functions introduced in 2012 by Beelen and Leander (these functions are not injective but we showed how any injective function G in $\frac{n}{2}$ variables can be used to make the parameterization F injective). Computer experiments provided some good results for the resulting Boolean functions, whose nonlinearity can be rather good and whose algebraic immunity was optimal for $n = 8, 10$, almost optimal for $n = 12, 14$, and a little worse for $n = 16$. Further work needs to be done, with adapted heuristics, for investigating other functions G (only a tiny part of them being possibly visited since their number is huge) for $n = 16$ and hopefully for $n = 18$. Other highly nonlinear injective vectorial functions will also have to be searched and used. Much work remains to do for studying more accurately the cryptographic parameters of parameterized functions, in particular their fast algebraic immunity, and for trying to reach still better nonlinearities and algebraic immunities with other choices of vectorial functions. We believe that parameterization opens an avenue for the design of Boolean functions to be used in cryptography.

Acknowledgement We deeply thank Stjepan Picek for his great help with experiments. We are indebted to him. We also thank Sihem Mesnager for her interesting indications.

References

[1] E. F. Assmus, Jr. and J. D. Key. *Designs and Their Codes*, volume 103 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, UK, 1992. See page 21.

[2] P. Beelen and G. Leander. A new construction of highly nonlinear S-boxes. *Cryptography and Communications* 4 (1), pp.65-77, 2012. See pages 3, 13, 14, and 20.

[3] A. Canteaut. Open problems related to algebraic attacks on stream ciphers. *Proceedings of Workshop on Coding and Cryptography WCC* 2005, pp. 1-10, 2005. See also a revised version in *Lecture Notes in Computer Science* 3969, pp. 120-134, 2006. See pages 11 and 25.

[4] C. Carlet. A larger Class of Cryptographic Boolean Functions via a Study of the Maiorana-McFarland Construction. *Proceedings of CRYPT0 2002, Lecture Notes in Computer Science* 2442, pp. 549-564, 2002. See page 22.

[5] C. Carlet. Handling vectorial functions by means of their graph indicators. *IEEE Transactions on Information Theory* 66 (10), pp. 6324-6339, 2020. See pages 8, 15, and 26.

[6] C. Carlet. Graph indicators of vectorial functions and bounds on the algebraic degree of composite functions. *IEEE Transactions on Information Theory* 66 (12), pp. 7702-7716, 2020. See pages 8, 15, and 26.

[7] C. Carlet. Boolean Functions for Cryptography and Coding Theory. Monograph in *Cambridge University Press*, 562 pages, 2021. See pages 2, 3, 4, 5, 6, 7, 8, 10, 11, 15, 16, 19, 20, 21, 22, and 28.

[8] C. Carlet, P. Charpin, and V. Zinoviev. Codes, bent functions and permutations suitable for DES-like cryptosystems. *Designs, Codes and Cryptography*, 15(2), pp. 125-156, 1998. See pages 8 and 13.

[9] C. Carlet, D. Dalai, K. Gupta and S. Maitra. Algebraic Immunity for Cryptographically Significant Boolean Functions: Analysis and Construction. *IEEE Transactions on Information Theory* 52 (7), pp. 3105-3121, 2006.

[10] C. Carlet and K. Feng. An infinite class of balanced functions with optimum algebraic immunity, good immunity to fast algebraic attacks and good nonlinearity. *Proceedings of ASIACRYPT 2008, Lecture Notes in Computer Science* 5350, pp. 425-440, 2008. See page 3.

[11] C. Carlet and P. Méaux. Boolean functions for homomorphic-friendly stream ciphers. *Proceedings of the Conference on Algebra, Codes and Cryptology (A2C)*, pp. 166-182, Springer, Cham 2019 (this version does not include proofs, a full paper will appear in IEEE Transactions on Information Theory). See page 3.

[12] C. Carlet and S. Picek. On the practical limits of a generalization of the hidden weight bit function and of another construction of highly nonlinear functions. Preprint, 2022. See pages 23 and 29.

[13] D. K. Dalai, S. Maitra and S. Sarkar. Basic Theory in Construction of Boolean Functions with Maximum Possible Annihilator Immunity. *Designs, Codes and Cryptography* 40 (1), pp. 41-58, 2006 (preliminary version available in *IACR Cryptology ePrint Archive* http://eprint.iacr.org/ 2005/229, 2005). See pages 16 and 21.

[14] C. Ding, C. Li and Y. Xia. Another generalisation of the binary Reed-Muller codes and its applications. *Finite Fields Their Appl.* 53, pp. 144-174, 2018. See page 21.

[15] H. Dobbertin. Construction of bent functions and balanced Boolean functions with high nonlinearity. *Proceedings of Fast Software Encryption FSE 1995, Lecture Notes in Computer Science* 1008, pp. 61-74, 1995. See page 19.

[16] I. Dumer and O. Kapralova. Spherically punctured reed-muller codes. *IEEE Transactions on Information Theory* 63 (5), pp. 2773-2780, 2017.

[17] K. Li, S. Mesnager and L. Qu. Further study of 2-to-1 mappings over $\mathbb{F}_2^n$. *IEEE Transactions on Information Theory* 67 (6), pp. 3486-3496, 2021. See page 10.

[18] F. J. MacWilliams and N. J. Sloane. *The theory of error-correcting codes*, Amsterdam, North Holland. 1977. See pages 5, 10, and 11.

[19] P. Méaux, C. Carlet, A. Journault and F.-X. Standaert. Improved Filter Permutators for Efficient FHE: Better Instances and Implementations. *Proceedings of Indocrypt 2019, Lecture Notes in Computer Science* 11898, pp. 68-91, 2019. See pages 2 and 3.

[20] P. Méaux, A. Journault, F.-X. Standaert and C. Carlet. Towards Stream Ciphers for Efficient FHE with Low-Noise Ciphertexts. *Proceedings of EUROCRYPT 2016, Lecture Notes in Computer Science* 9665, pp. 311-343, 2016. See pages 2 and 3.

[21] S. Mesnager. "Linear codes from functions", Chapter 20 in "A Concise Encyclopedia 1419 Coding Theory" CRC Press/Taylor and Francis Group (Publisher), London, New York, 2021 (94 pages). See page 21.

[22] S. Mesnager, L. Qu. On Two-to-One Mappings Over Finite Fields. *IEEE Transactions on Information Theory* 65 (12), pp. 7884-7895, 2019. See page 10.

[23] G. Mullen and D. Panario. *Handbook of Finite Fields*. CRC Press Book, 2013. See page 24.

[24] Q. Wang, C. Carlet, P. Stănică and C. H. Tan. Cryptographic Properties of the Hidden Weighted Bit Function. *Discrete Applied Mathematics* 174, pp. 1-10, 2014. See page 25.