Weightwise perfectly balanced functions and nonlinearity

Agnese Gini, Pierrick Méaux

University of Luxembourg, Luxembourg
agnese.gini@uni.lu, pierrick.meaux@uni.lu

Abstract. In this article we realize a general study on the nonlinearity of weightwise perfectly balanced (WPB) functions. First, we derive upper and lower bounds on the nonlinearity from this class of functions for all n. Then, we give a general construction that allows us to provably provide WPB functions with nonlinearity as low as $2^{n/2-1}$ and WPB functions with high nonlinearity, at least $2^{n-1} - 2^{n/2}$. We provide concrete examples in 8 and 16 variables with high nonlinearity given by this construction. In 8 variables we experimentally obtain functions reaching a nonlinearity of 116 which corresponds to the upper bound of Dobbertin's conjecture, and it improves upon the maximal nonlinearity of WPB functions recently obtained with genetic algorithms. Finally, we study the distribution of nonlinearity over the set of WPB functions. We examine the exact distribution for $n = 4$ and provide an algorithm to estimate the distributions for $n = 8$ and 16, together with the results of our experimental studies for $n = 8$ and 16.

1 Introduction.

Boolean functions have multiple applications in secure communications, therefore numerous criteria to determine suitable functions for the specific applications have been proposed during the last decades. In 2017, Carlet, Méaux, and Rotella began to study the cryptographic criteria of Boolean functions with restricted input set [CMR17], motivated by the cryptanalyzis of the FLIP stream cipher [MJSC16], introduced for hybrid homomorphic encryption. FLIP's peculiarity is that its filter function is evaluated on sets of Boolean vectors having constant Hamming weight. Thus, having functions with good properties also when restricted is crucial for examining its security. For instance, as generally working with balanced functions avoids biased output distributions, it is preferable for applications like FLIP to work with functions balanced when restricted over the slices $\mathsf{E}_{k,n} = \{x \in \mathbb{F}_2^n \,|\, \mathsf{w}_\mathsf{H}(x) = k\}$ of the hypercube $\mathbb{F}_2^n$. To study this case, Carlet *et al.* [CMR17] introduced the notion of Weightwise Perfectly Balanced (WPB) functions, *i.e.* $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$, such that $|\{x \in \mathsf{E}_{k,n} \,|\, f(x) = 0\}| = |\{x \in \mathsf{E}_{k,n} \,|\, f(x) = 1\}|$ for each $1 \leq k \leq n-1$, and $f(0) = 0$ and $f(1) = 1$. The authors observed that these functions exist only when n is a power of two $n = 2^m$, and provided explicit constructions. Since then, many methods for constructing WPB functions have been proposed; *e.g.* [LM19, TL19, LS20, MS21, ZS21, MSL21, GS22, ZS22, MPJ+22, GM22a, GM22b, MKCL22, MSLZ22].

In particular, Mandujano, Ku Cauich, and Lara recently investigated the problem of finding WPB functions with high nonlinearity via genetic algorithms [MKCL22]. While Boolean functions having the highest possible nonlinearity (*i.e.* bent functions) are known for $n = 2^m$, the problem of determining the maximal nonlinearity for a balanced 2^m-variable function is still open, and *a fortiori* for a WPB function. Indeed, being balanced, WPB functions cannot be bent (see *e.g.* [SZZ93, Car21]). In general, evolutionary algorithms give quite good results for small value of n [MCD97, PCG+16, MPJ+22], and in fact the authors of [MKCL22] were able to find a 8-variable function with nonlinearity 112, which was the maximal obtained so far for a WPB function. However, for larger values of n their results are limited due to the massive computational power required for this kind of approach.

Hence, the goal of this paper is to further investigate the nonlinearity of WPB functions, from a more algebraic point of view. Namely, we first discuss the known upper bound and determine novel lower bounds on the nonlinearity of a WPB function. For this purpose we introduce the notion of Non Perfect Balancedness (NPB), which measures the distance of Boolean function from the family of WPB functions. This notion is the analogous of the nonlinearity respect to affine functions. Then, we address the problem of building WPB functions with prescribed nonlinearity. Specifically, our construction modifies the support of the input function on each slice to make it perfectly balanced, so that the output is a WPB function lying at minimal Hamming distance from the input function. By using this construction, we are able to exhibit both WPB functions with very low and very high nonlinearity for any n. Thereafter, we instantiate this strategy for 8 and 16 variables, obtaining two large families of WPB functions with

almost optimal nonlinearity. For instance, we prove that for $n = 8$ our construction gives more than 2^{43} different functions with nonlinearity at least 112. We also provide explicit examples with nonlinearity $112, 114$ and 116, which is therefore the new highest value known (and reaching the highest nonlinearity observed for a 8-variable balanced functions). Finally, we study the nonlinearity distribution of WPB functions. As in [MCD97] the results obtained via evolutionary algorithm are compared with uniform distribution, in this paper we analyze the nonlinearity's behavior of WPB functions sampled uniformly at random. We followed the model establish by [GM22a] for the weightwise nonlinearities of WPB functions. As a result of our experiments, we observed that extreme values are rare, concluding that we cannot expect in practice to find large families of functions, like those we exhibit from our construction, by sampling uniformly at random.

2 Preliminaries

For readability, we use the notation $+$ instead of $\oplus$ to denote the addition in $\mathbb{F}_2$ and $\sum$ instead of $\bigoplus$. In addition to classic notations we use $[a, b]$ to denote the subset of all integers between a and b: $\{a, a + 1, \ldots, b\}$. For a vector $v \in \mathbb{F}_2^n$ we denote $\mathsf{w_H}(v)$ its Hamming weight $\mathsf{w_H}(v) = |\{i \in [1, n] \,|\, v_i = 1\}|$. For two vectors v and w of $\mathbb{F}_2^n$ we denote $\mathsf{d_H}(v, w)$ the Hamming distance between v and w, $\mathsf{d_H}(v, w) = \mathsf{w_H}(v + w)$.

2.1 Boolean functions and weightwise considerations

In this part we recall the main concepts on Boolean functions used in cryptography and their weightwise properties we will use in this article. We refer to *e.g.* [Car21] for Boolean functions and cryptographic parameters and to [CMR17] for the weightwise properties, also named as properties on the slices. For $k \in [0, n]$ we call slice of the Boolean hypercube (of dimension n) the set $\mathsf{E}_{k,n} = \{x \in \mathbb{F}_2^n \,|\, \mathsf{w_H}(x) = k\}$. Thereafter, the Boolean hypercube is partitioned into $n + 1$ slices where the elements have the same Hamming weight.

Definition 1 (Boolean Function). *A Boolean function f in n variables is a function from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. The set of all Boolean functions in n variables is denoted by $\mathcal{B}_n$.*

Definition 2 (Algebraic Normal Form (ANF) and degree). *We call Algebraic Normal Form of a Boolean function f its n-variable polynomial representation over $\mathbb{F}_2$ (i.e. belonging to $\mathbb{F}_2[x_1, \ldots, x_n]/(x_1^2 + x_1, \ldots, x_n^2 + x_n)$):*

$$f(x_1, \ldots, x_n) = \sum_{I \subseteq [1,n]} a_I \left(\prod_{i \in I} x_i \right)$$

where $a_I \in \mathbb{F}_2$. The (algebraic) degree of f, denoted $\mathsf{deg}(f)$ is:

$$\mathsf{deg}(f) = \max_{I \subseteq [1,n]} \{|I| \,|\, a_I = 1\} \text{ if } f \text{ is not null}, 0 \text{ otherwise}.$$

To denote when a property or a definition is restricted to a slice we use the subscript k. For example, for a n-variable Boolean function f we denote its support $\mathsf{supp}(f) = \{x \in \mathbb{F}_2^n \,|\, f(x) = 1\}$ and we refer to $\mathsf{supp}_k(f)$ for its support restricted to a slice, *i.e.* $\mathsf{supp}(f) \cap \mathsf{E}_{k,n}$.

Definition 3 (Balancedness). *A Boolean function $f \in \mathcal{B}_n$ is called balanced if $|\mathsf{supp}(f)| = 2^{n-1} = |\mathsf{supp}(f + 1)|$.*

For $k \in [0, n]$ the function is said balanced on the slice k if $||\mathsf{supp}_k(f)| - |\mathsf{supp}_k(f + 1)|| \leq 1$. In particular when $|\mathsf{E}_{k,n}|$ is even $|\mathsf{supp}_k(f)| = |\mathsf{supp}_k(f + 1)| = |\mathsf{E}_{k,n}|/2$.

Definition 4 (Nonlinearity and weightwise nonlinearity). *The nonlinearity $\mathsf{NL}(f)$ of a Boolean function $f \in \mathcal{B}_n$, where n is a positive integer, is the minimum Hamming distance between f and all the affine functions in $\mathcal{B}_n$:*

$$\mathsf{NL}(f) = \min_{g,\, \mathsf{deg}(g) \leq 1} \{d_H(f, g)\},$$

where $g(x) = a \cdot x + \varepsilon$, $a \in \mathbb{F}_2^n, \varepsilon \in \mathbb{F}_2$ (where $\cdot$ is an inner product in $\mathbb{F}_2^n$, any choice of inner product will give the same value of $\mathsf{NL}(f)$).

For $k \in [0,n]$ we denote NL_k the nonlinearity on the slice k, the minimum Hamming distance between f restricted to $\mathsf{E}_{k,n}$ and the restrictions to $\mathsf{E}_{k,n}$ of affine functions over $\mathbb{F}_2^n$. Accordingly:

$$\mathsf{NL}_k(f) = \min_{g,\, \deg(g)\leq 1} |\mathsf{supp}_k(f+g)|.$$

We refer to the global weightwise nonlinearity of f as $\mathsf{GWNL}(f) = \sum_{k=0}^{n} \mathsf{NL}_k(f)$.

The functions reaching the maximal value of nonlinearity are called *bent*, and are deeply studied in the context of symmetric cryptography (see *e.g.* [Rot76, Tok15, Mes16]).

Definition 5 (Bent function). *Let $n \in \mathbb{N}^*$ be even. A Boolean function $f \in \mathcal{B}_n$ is* bent *if and only if $\mathsf{NL}(f) = 2^{n-1} - 2^{n/2-1}$.*

We also recall the concept of Walsh transform, and restricted Walsh transform [MMM+18], which are of particular interest to study the (restricted) nonlinearity or balancedness.

Definition 6 (Walsh transform and restricted Walsh transform). *Let $f \in \mathcal{B}_n$ be a Boolean function, its Walsh transform W_f at $a \in \mathbb{F}_2^n$ is defined as:*

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

Let $f \in \mathcal{B}_n$, $S \subset \mathbb{F}_2^n$, its Walsh transform restricted to S at $a \in \mathbb{F}_2^n$ is defined as:

$$W_{f,S}(a) := \sum_{x \in S} (-1)^{f(x)+ax}.$$

For $S = \mathsf{E}_{k,n}$ we denote $W_{f,\mathsf{E}_{k,n}}(a)$ by $\mathcal{W}_{f,k}(a)$, and for $a = 0_n$ we denote $\mathcal{W}_{f,k}(a)$ as $\mathcal{W}_{f,k}(\mathbf{0})$.

Property 1 (Nonlinearity and Walsh transform, *e.g.* [Car21])**.** *Let $n \in \mathbb{N}^*$, for every n-variable Boolean function f:*

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

Property 2 (Nonlinearity on the slice and restricted Walsh transform, adapted from [CMR17], Proposition 6)**.** *Let $n \in \mathbb{N}^*, k \in [0,n]$, for every n-variable Boolean function f over $\mathsf{E}_{k,n}$:*

$$\mathsf{NL}_k(f) = \frac{|\mathsf{E}_{k,n}|}{2} - \frac{\max_{a \in \mathbb{F}_2^n} |\mathcal{W}_{f,k}(a)|}{2}.$$

Property 3 (Balancedness on the slice and restricted Walsh transform [GM22b])**.** *Let $n \in \mathbb{N}^*, k \in [0,n]$, $f \in \mathcal{B}_n$ is balanced over $\mathsf{E}_{k,n}$ if and only if:*

$$\mathcal{W}_{f,k}(0_{|\mathsf{E}_{k,n}|}) = \begin{cases} 0 & \text{if } |\mathsf{E}_{k,n}| \text{ is even,} \\ \pm 1 & \text{if } |\mathsf{E}_{k,n}| \text{ is odd.} \end{cases}$$

2.2 Symmetric Functions

The n-variable *Boolean symmetric functions* are those that are constant on each slice $\mathsf{E}_{k,n}$ for $k \in [0,n]$. This class has been assiduously studied in the context of cryptography, see *e.g.* [Car04, CV05, BP05, SM07, QFLW09, Méa21, CM21]. In this paper we mainly consider two families of symmetric functions, which are both bases of the symmetric functions:

Definition 7 (Elementary symmetric functions). *Let $i \in [0,n]$, the elementary symmetric function of degree i in n variables, denoted $\sigma_{i,n}$, is the function which ANF contains all monomials of degree i and no monomials of other degrees.*

Definition 8 (Slice indicator functions). *Let $k \in [0, n]$, the indicator function of the slice of weight k is defined as:*

$$\forall x \in \mathbb{F}_2^n, \quad \varphi_{k,n}(x) = 1 \text{ if and only if } \mathsf{w_H}(x) = k.$$

Property 4 (Nonlinearity of $\sigma_{2,n}$)**.** *Let $n \in \mathbb{N}^*$ even, the elementary symmetric function $\sigma_{2,n} = \sum_{1 \le i < j \le n} x_i x_j$ is bent, i.e. $\mathsf{NL}(\sigma_{2,n}) = 2^{n-1} - 2^{n/2-1}$.*

Property 5 (Weightwise restricted Walsh transform and addition of symmetric function ([GM22b], Proposition 4))**.** *Let $n \in \mathbb{N}^*$, $k \in [0, n]$ and $f \in \mathcal{B}_n$, the following holds on $f + \varphi_{k,n}$*

$$\forall a \in \mathbb{F}_2^n, \forall i \in [0, n] \setminus \{k\}, \mathcal{W}_{f+\varphi_{k,n},i}(a) = \mathcal{W}_{f,i}(a), \text{ and } \mathcal{W}_{f+\varphi_{k,n},k}(a) = -\mathcal{W}_{f,i}(a).$$

2.3 Weightwise perfectly balanced functions

Definition 9 (Weightwise Perfectly Balanced Function (WPB)). *Let $m \in \mathbb{N}^*$ and f be a Boolean function in $n = 2^m$ variables. It will be called weightwise perfectly balanced (WPB) if, for every $k \in [1, n-1]$, f is balanced on the slice k, that is $\forall k \in [1, n-1], |\mathsf{supp}_k(f)| = \binom{n}{k}/2$, and:*

$$f(0, \ldots, 0) = 0, \quad \text{and } f(1, \ldots, 1) = 1.$$

The set of WPB functions in 2^m variables is denoted $\mathcal{WPB}_m$.

Property 6 (WPB functions, alternative definition)**.** *Let $m \in \mathbb{N}^*$, $n = 2^m$, f is a WPB function if:*

- $f(0_n) = 0$,
- $\forall k \in [1, n-1]$, $\mathcal{W}_{f,k}(\mathbf{0}) = 0$,
- $f(1_n) = 1$.

2.4 Krawtchouk polynomials and properties

For some proofs we will use Krawtchouk polynomials and some of their properties, we give the necessary preliminaries here and refer to *e.g.* [MS78] for more details. ¨

Definition 10 (Krawtchouk Polynomials). *The Krawtchouk polynomial of degree k, with $0 \le k \le n$ is given by:* $\mathsf{K}_k(\ell, n) = \sum_{j=0}^{k} (-1)^j \binom{\ell}{j} \binom{n-\ell}{k-j}$.

Property 7 (Krawtchouk polynomials relations)**.** *Let $n \in \mathbb{N}^*$ and $k \in [0, n]$, the following relations hold:*

1. $\mathsf{K}_k(\ell, n) = \sum_{x \in \mathsf{E}_{k,n}} (-1)^{a \cdot x}$, *where $a \in \mathbb{F}_2^n$ and $\ell = \mathsf{w_H}(a)$,*
2. $\mathsf{K}_{n-k}(\ell, n) = (-1)^\ell \mathsf{K}_k(\ell, n)$,
3. $\mathsf{K}_k(n - \ell, n) = (-1)^k \mathsf{K}_k(\ell, n)$,

Property 8 (Proposition 5 [DMS06])**.** *For n even, $k \in [0, n]$,*

$$\mathsf{K}_k(n/2, n) = \begin{cases} 0 & \text{if } k \text{ is odd}, \\ (-1)^{k/2} \binom{n/2}{k/2} & \text{if } k \text{ is even}. \end{cases}$$

3 Bounds on the nonlinearity of a WPB function

For n-variable Boolean functions the upper bound on the nonlinearity is a classic result, it is $2^{n-1} - 2^{n/2-1}$ and it can be reached only for n even, by bent functions. Since WPB functions are balanced, they cannot be bent (see *e.g.* [SZZ93, Car21]), therefore no WPB function can reach this bound, and we will consider as upper bound the one holding for all balanced functions. Therefore, from [SZZ93] Corollary 7, we have that a WPB function f has nonlinearity at most:

$$U_m = 2^{n-1} - 2^{n/2-1} - 2, \tag{1}$$

since $n = 2^m$ is even and f balanced. In this section we will focus on the lower bound on the nonlinearity of WPB functions, and provide different lower bounds.

First, we derive a lower bound based on the results on the weightwise nonlinearities of WPB functions.

Proposition 1. *Let $m \in \mathbb{N}^*$, $n = 2^m$ and $f \in \mathcal{WPB}_m$, then:*

$$\mu_n \leq \mathsf{GWNL}(f) \leq \mathsf{NL}(f),$$

where μ_n is the global minimum weightwise nonlinearity $\mu_n = \sum_{k=1}^{n-1} \min_{g \in \mathcal{WPB}_m} \mathsf{NL}_k(g)$.

Proof. The first inequality comes from the definition of GWNL (see Definition 4), and the second one comes from the fact that the nonlinearity considers the best affine approximation over $\mathbb{F}_2^n$ whereas the global weightwise nonlinearity considers the best affine approximation on each slice. □

Thereafter, lower bounds on μ_n allow us to derive lower bounds on $\mathsf{NL}(f)$ when f is WPB. Using the results from [GM22a], Proposition 7, it gives:

$$\mathsf{NL}(f) \geq \begin{cases} 2 & \text{if } m = 3, \\ 4 & \text{if } m > 3, m \text{ even}, \\ 6 & \text{if } m > 3, m \text{ odd}. \end{cases}$$

To improve upon this bound we consider the distance between affine functions and WPB functions. To do so we introduce the notion of *Non Perfect Balancedness (NPB)*, similar to the nonlinearity.

Definition 11 (Non Perfect Balancedness). *Let $m \in \mathbb{N}^*$, $n = 2^m$, and f an n-variable Boolean function, the non perfect balancedness of f, denoted $\mathsf{NPB}(f)$ is defined as:*

$$\mathsf{NPB}(f) = \min_{g \in \mathcal{WPB}_m} \mathsf{d_H}(f, g).$$

The NPB measures the distance to WPB functions such as the nonlinearity measures the distance to affine functions. In the following we give an expression of the NPB from the restricted Walsh transform.

Proposition 2 (NPB and restricted Walsh transform). *Let $m \in \mathbb{N}^*$, $n = 2^m$, and $f \in \mathcal{B}_n$, the following holds on its non perfect balancedness:*

$$\mathsf{NPB}(f) = \frac{2 - \mathcal{W}_{f,0}(\mathbf{0}) + \mathcal{W}_{f,n}(\mathbf{0})}{2} + \sum_{k=1}^{n-1} \frac{|\mathcal{W}_{f,k}(\mathbf{0})|}{2}.$$

Proof. First, we rewrite the expression of NPB by partitioning $\mathbb{F}_2^n$ into the $n+1$ slices. For h a Boolean function we denote $\mathsf{supp}_k(h)$ the support of h on $\mathsf{E}_{k,n}$ and h_k the restriction of h on $\mathsf{E}_{k,n}$. Moreover, denoting by ν_k the binomial

$\binom{n}{k}$ we have

$$\mathsf{NPB}(f) = \min_{g \in \mathcal{WPB}_m} \mathsf{d_H}(f,g) = \min_{g \in \mathcal{WPB}_m} \sum_{k=0}^{n} \mathsf{d_H}(f_k, g_k) \tag{2}$$

$$= |\mathsf{supp}_0(f)| + 1 - |\mathsf{supp}_n(f)| + \sum_{k=1}^{n-1} \min_{v \in \mathsf{E}_{\frac{\nu_k}{2},\nu_k}} \mathsf{d_H}(f_k, v) \tag{3}$$

$$= |\mathsf{supp}_0(f)| + 1 - |\mathsf{supp}_n(f)| + \sum_{k=1}^{n-1} |\frac{\nu_k}{2} - |\mathsf{supp}_k(f)|| \tag{4}$$

$$= \frac{1 - \mathcal{W}_{f,0}(\mathbf{0})}{2} + \frac{1 + \mathcal{W}_{f,n}(\mathbf{0})}{2} + \sum_{k=1}^{n-1} \frac{|\mathcal{W}_{f,k}(\mathbf{0})|}{2}, \tag{5}$$

where Equation 3 is obtained by using the definition of WPB functions (Definition 9), *i.e.* their value is 0 in 0_n and 1 in 1_n, and from the fact that on the slice $k \in [1, n-1]$ each element of Hamming weight $\nu_k/2$ is the support of WPB functions. Finally, Equation 5 comes from the expression of $\mathcal{W}_{f,k}(\mathbf{0})$:

$$\mathcal{W}_{f,k}(\mathbf{0}) = \sum_{x \in \mathsf{E}_{k,n}} (-1)^{f(x)} = \nu_k - 2|\mathsf{supp}_k(f)|.$$

□

Then, we can express the non perfect balancedness of affine functions in terms of sum of Krawtchouk polynomials, which will be an important step to derive the final lower bound.

Lemma 1 (NPB of affine functions). *Let $m \in \mathbb{N}^*$, $n = 2^m$, and $f \in \mathcal{B}_n$ be an affine function. Let us denote $f = ax + \varepsilon$ with $a \in \mathsf{E}_{\ell,n}$ and $\varepsilon \in \{0,1\}$, its non perfect balancedness is:*

$$\mathsf{NPB}(f) = \varepsilon + (\ell + 1 + \varepsilon \mod 2) + \sum_{k=1}^{n/2-1} |\mathsf{K}_k(\ell, n)| + \frac{|\mathsf{K}_{\frac{n}{2}}(\ell, n)|}{2}.$$

Proof. First, we give the relation between $\mathcal{W}_{f,k}(\mathbf{0})$ and $\mathsf{K}_k(\ell, n)$ when $f = ax + \varepsilon$, using the first item of Property 7. For $k \in [0, n]$:

$$\mathcal{W}_{f,k}(\mathbf{0}) = \sum_{x \in \mathsf{E}_{k,n}} (-1)^{ax+\varepsilon} = (-1)^{\varepsilon} \sum_{x \in \mathsf{E}_{k,n}} (-1)^{ax} = (-1)^{\varepsilon} \mathsf{K}_k(\ell, n).$$

Then, using Proposition 2 we get:

$$\mathsf{NPB}(f) = \frac{\sum_{k=1}^{n-1} |(-1)^{\varepsilon} \mathsf{K}_k(\ell, n)|}{2} + \frac{2 - (-1)^{\varepsilon} \mathsf{K}_0(\ell, n) + (-1)^{\varepsilon} \mathsf{K}_n(\ell, n)}{2} \tag{6}$$

$$= \sum_{k=1}^{n/2-1} |\mathsf{K}_k(\ell, n)| + \frac{|\mathsf{K}_{\frac{n}{2}}(\ell, n)|}{2} + \frac{2 + ((-1)^{\varepsilon+1} + (-1)^{\varepsilon+\ell}) \mathsf{K}_0(\ell, n)}{2} \tag{7}$$

$$= \sum_{k=1}^{n/2-1} |\mathsf{K}_k(\ell, n)| + \frac{|\mathsf{K}_{\frac{n}{2}}(\ell, n)|}{2} + 1 + \frac{(-1)^{\varepsilon+1} + (-1)^{\varepsilon+\ell}}{2}. \tag{8}$$

Equation 7 is obtained using Property 7 Item 2, and Equation 8 comes from $\mathsf{K}_0(\ell, n) = 1$ using Definition 10. Finally, we rewrite $1 + \frac{(-1)^{\varepsilon+1} + (-1)^{\varepsilon+\ell}}{2}$ as $\varepsilon + (\ell + 1 + \varepsilon \mod 2)$, which can be verified considering the four cases:

- ℓ odd and $\varepsilon = 0$ giving 0,
- ℓ odd and $\varepsilon = 1$ giving 2,

– ℓ even and $\varepsilon = 0$ giving 1,
– ℓ even and $\varepsilon = 1$ giving 1.

□

We highlight the NPB of particular affine functions, the linear functions having 0, 1 and $n/2$ monomials in their ANF.

Proposition 3. *Let $m \in \mathbb{N}^*$ and $n = 2^m$, we consider the three following n-variable Boolean functions $f(x) = 0$, $g(x) = x_1$ and $h(x) = \sum_{i=1}^{n/2} x_i$. Their non perfect balancedness is the following:*

– $\mathsf{NPB}(f) = 2^{n-1}$,
– $\mathsf{NPB}(g) = \binom{n-1}{n/2-1} - 1$,
– $\mathsf{NPB}(h) = 2^{n/2-1}$.

Proof. We begin with the expression of $\mathsf{NPB}(f)$ using Lemma 1. For $k \in [0, n]$ the Krawtchouk polynomial $\mathsf{K}_k(0, n)$ takes the value $\binom{n}{k}$, accordingly:

$$\mathsf{NPB}(f) = 1 + \sum_{k=1}^{n/2-1} |\binom{n}{k}| + \frac{\binom{n}{n/2}}{2} = \frac{1}{2}\sum_{k=0}^{n} \binom{n}{k} = 2^{n-1}.$$

Then, we give the expression of $\mathsf{NPB}(g)$ (using Lemma 1), using that for $k \in [0, n]$ $\mathsf{K}_k(1, n) = \binom{n-1}{k} - \binom{n-1}{k-1}$. Since for $k \in [1, n/2 - 1]$ we have $\binom{n-1}{k} \geq \binom{n-1}{k-1}$, $\mathsf{K}_k(1, n)$ is positive. Thereafter,

$$\begin{aligned}\mathsf{NPB}(g) &= \sum_{k=1}^{n/2-1} \binom{n-1}{k} - \binom{n-1}{k-1} + \frac{|\binom{n-1}{n/2} - \binom{n-1}{n/2-1}|}{2} \\ &= \binom{n-1}{n/2-1} - \binom{n-1}{0} + 0 = \binom{n-1}{n/2-1} - 1.\end{aligned}$$

Finally, we give the expression of $\mathsf{NPB}(h)$ using Lemma 1, Property 7 Item 2, and Property 8 for the value of Krawtchouk polynomials:

$$\begin{aligned}\mathsf{NPB}(h) &= 1 + \sum_{k=1}^{n/2-1} |\mathsf{K}_k(n/2, n)| + \frac{|\mathsf{K}_{\frac{n}{2}}(n/2, n)|}{2} = \frac{1}{2}\sum_{k=0}^{n} |\mathsf{K}_k(n/2, n)| \\ &= \frac{1}{2}\sum_{t=0}^{n/2} |(-1)^{2t/2}\binom{n/2}{2t/2}| = \frac{1}{2}\sum_{t=0}^{n/2} \binom{n/2}{t} = 2^{n/2-1}.\end{aligned}$$

□

Finally, we provide a lower bound on the nonlinearity of WPB functions using the NPB of affine functions.

Theorem 1 (Lower bound on the nonlinearity of WPB functions). *Let $m \in \mathbb{N}$, $m \geq 2$, $n = 2^m$, and B_m be the integer defined as:*

$$B_m = \min_{\substack{\ell \in [0, n/2] \\ \varepsilon \in \{0,1\}}} \varepsilon + (\ell + 1 + \varepsilon \mod 2) + \sum_{k=1}^{n/2-1} |\mathsf{K}_k(\ell, n)| + \frac{|\mathsf{K}_{\frac{n}{2}}(\ell, n)|}{2},$$

then, $\forall f \in \mathcal{WPB}_m$, $\mathsf{NL}(f) \geq B_m$.

Proof. Using Property 7 Item 3 and Lemma 1 we obtain that B_m is the NPB minimal over the n-variable affine functions. Then, using the definitions of nonlinearity and non perfect balancedness, for f WPB we obtain:

$$\mathsf{NL}(f) = \min_{g \text{ affine}} \mathsf{d_H}(f,g) \geq \min_{\substack{g \text{ affine} \\ f \in \mathcal{WPB}_m}} \mathsf{d_H}(f,g) = \min_{g \text{ affine}} \mathsf{NPB}(g) = B_m.$$

□

Remark 1. We computed explicitly B_m for small value of m; see Table 1. From Proposition 3 we get that B_m is at least $2^{n/2-1}$, *i.e.* the non perfect balancedness given by a linear function with $\frac{n}{2}$ monomials in its ANF, *e.g.* $h(x) = \sum_{i=1}^{n/2} x_i$. Table 1 shows that actually for m up to 6, *i.e.* 64 variables, $B_m = \mathsf{NPB}(h)$.

m	2	3	4	5	6
B_m	2	8	128	2^{15}	2^{31}
U_m	4	118	32638	$2^{31} - 2^{15} - 2$	$2^{63} - 2^{31} - 2$

Table 1. Concrete values of B_m and U_m for small values of m.

4 Constructions of WPB functions with prescribed nonlinearity

In this section we present a construction allowing to obtain a WPB function from any 2^m-variable Boolean function f. The principle of the construction is to modify the support of the input function on each slice to make it perfectly balanced, enabling us to obtain as output a WPB function g which lies at distance $\mathsf{NPB}(f)$ from the input function. We show thereafter how we can use this construction to build functions with low, or high nonlinearity.

Construction 1

Input: Let $m \in \mathbb{N}$, $m \geq 2$, $n = 2^m$ and f a n-variable function.
Output: $g \in \mathcal{WPB}_m$.

1: Initiate the support of g to $\mathsf{supp}(f)$.
2: If $0_n \in \mathsf{supp}(f)$ remove 0_n from $\mathsf{supp}(g)$.
3: If $1_n \notin \mathsf{supp}(f)$ add 1_n to $\mathsf{supp}(g)$.
4: **for** $k \leftarrow 1$ to $n-1$ **do**
5: Compute $C_{k,n} = \mathcal{W}_{f,k}(\mathbf{0})/2$,
6: **if** $C_{k,n} < 0$ **then**
7: remove $|C_{k,n}|$ elements from $\mathsf{supp}_k(g)$,
8: **else**
9: **if** $C_{k,n} > 0$ **then**
10: add $C_{k,n}$ new elements to $\mathsf{supp}_k(g)$,
11: **end if**
12: **end if**
13: **end for**
14: **return** g

Theorem 2 (Weightwise perfect balancedness and distance of Construction 1). *Let $m \in \mathbb{N}$, $m \geq 2$ and $n = 2^m$. Any function given by Construction 1 with input f is weightwise perfectly balanced, and $\mathsf{d_H}(f,g) = \mathsf{NPB}(f)$.*

Proof. First, we show that g is WPB, using the characterization from Property 6:

- $g(0_n) = 0$ since 0_n does not belong to $\mathsf{supp}(g)$.
- $g(1_n) = 1$ since 1_n belongs to $\mathsf{supp}(g)$.
- For $k \in [1, n-1]$, by construction if $C_{k,n}$ is inferior to zero then $\mathsf{supp}_k(g)$ has $|C_{k,n}|$ elements less than $\mathsf{supp}_k(f)$. We study what it implies on the restricted Walsh transform of g:

$$\begin{aligned}\mathcal{W}_{g,k}(\mathbf{0}) &= \sum_{x \in \mathsf{E}_{k,n}} (-1)^{g(x)} = |x \in \{\mathsf{E}_{k,n} \setminus \mathsf{supp}_k(g)\}| - |x \in \mathsf{supp}_k(g)| \\ &= |x \in \{\mathsf{E}_{k,n} \setminus \mathsf{supp}_k(f)\}| + |C_{k,n}| - (|x \in \mathsf{supp}_k(f)| - |C_{k,n}|) \\ &= \mathcal{W}_{f,k}(\mathbf{0}) + 2|C_{k,n}| = 0.\end{aligned}$$

  In the other case, if $C_{k,n} = 0$, f is already balanced on the slice k then g is equal to f on this slice and $\mathcal{W}_{g,k}(\mathbf{0}) = \mathcal{W}_{f,k}(\mathbf{0}) = 0$. If $C_{k,n} > 0$ by construction $\mathsf{supp}_k(g)$ has $C_{k,n}$ elements more than $\mathsf{supp}_k(f)$, and similarly to the case $C_{k,n} < 0$ we obtain:

$$\begin{aligned}\mathcal{W}_{g,k}(\mathbf{0}) &= |x \in \{\mathsf{E}_{k,n} \setminus \mathsf{supp}_k(f)\}| - |C_{k,n}| - (|x \in \mathsf{supp}_k(f)| + |C_{k,n}|) \\ &= \mathcal{W}_{f,k}(\mathbf{0}) - 2|C_{k,n}| = 0.\end{aligned}$$

  It allows us to conclude that $g \in \mathcal{WPB}_m$.

  Then, we show that $\mathsf{d_H}(f, g) = \mathsf{NPB}(f)$. To do so we study the distance between f and g on each slice, denoting $\mathsf{d_{H,k}}(f, g) = |\{x \in \mathsf{E}_{k,n} \text{ such that } f(x) \neq g(x)\}|$.

- For $k = 0$, $\mathsf{supp}_0(g)$ is forced to be $\emptyset$, therefore $\mathsf{d_{H,0}}(f, g) = 0$ if $\mathsf{supp}_0(f) = \emptyset$ and $\mathsf{d_{H,0}}(f, g) = 1$ otherwise, which is equivalent to $\mathsf{d_{H,0}}(f, g) = (1 - \mathcal{W}_{f,0}(\mathbf{0}))/2$.
- For $k = n$, $\mathsf{supp}_n(g)$ is forced to be $\{1_n\}$, therefore $\mathsf{d_{H,0}}(f, g) = 1$ if $\mathsf{supp}_n(f) = \emptyset$ and $\mathsf{d_{H,n}}(f, g) = 0$ otherwise, which is equivalent to $\mathsf{d_{H,n}}(f, g) = (1 + \mathcal{W}_{f,n}(\mathbf{0}))/2$.
- For $k \in [1, n-1]$, $|C_{k,n}|$ elements are removed or added to $\mathsf{supp}_k(g)$ hence $\mathsf{d_{H,k}}(f, g) = |C_{k,n}| = |\mathcal{W}_{f,k}(\mathbf{0})|/2$.

Summing over all $k \in [0, n]$ and using Proposition 2 we can conclude:

$$\mathsf{d_H}(f, g) = \frac{2 - \mathcal{W}_{f,0}(\mathbf{0}) + \mathcal{W}_{f,n}(\mathbf{0})}{2} + \sum_{k=1}^{n-1} \frac{|\mathcal{W}_{f,k}(\mathbf{0})|}{2} = \mathsf{NPB}(f).$$

□

As a first application we show how to obtain WPB functions with very low nonlinearity.

Proposition 4 (WPB function with low nonlinearity). *Let $m \in \mathbb{N}$, $m \geq 2$ and $n = 2^m$, there exists WPB functions g such that $\mathsf{NL}(g) = 2^{n/2-1}$.*

Proof. We prove the existence by exhibiting such functions. We define f as $\sum_{i=1}^{n/2} x_i$, such function is linear, and using Proposition 3 we have $\mathsf{NPB}(f) = 2^{n/2-1}$. Accordingly, using Construction 1 seeded with f, we obtain a function g WPB such that $\mathsf{d_H}(f, g) = 2^{n/2-1}$. Since the distance between two n-variable affine functions is at least 2^{n-1} (minimal distance of order-1 Reed-Muller code), f is the affine function the closest to g if $2^{n/2-1} \leq 2^{n-2}$, that is if $n \geq 2$. Thereafter, $\mathsf{NL}(g) = \mathsf{d_H}(f, g) = 2^{n/2-1}$.

□

Proposition 5 (WPB function with high nonlinearity). *Let $m \in \mathbb{N}$, $m \geq 2$ and $n = 2^m$. Consider $f_n = \sigma_{2,n} + \ell_{n/2}$ where $\ell_{n/2} = \sum_{i=1}^{n/2} x_i$. Construction 1 applied with f_n as input returns a WPB function g such that $\mathsf{NL}(g) \geq 2^{n-1} - 2^{n/2}$.*

Proof. Since $\sigma_{2,n}$ is a symmetric function giving 0 in 0_n and 1_n, it can be decomposed as a sum of $\varphi_{k,n}$ with $k \in [1, n-1]$, therefore using Property 5 we have:

- $\mathcal{W}_{f_n,0}(\mathbf{0}) = \mathcal{W}_{\ell_{n/2},0}(\mathbf{0})$,
- for $k \in [1, n-1]$, $\mathcal{W}_{f_n,0}(\mathbf{0}) = \pm\mathcal{W}_{\ell_{n/2},0}(\mathbf{0})$,
- $\mathcal{W}_{f_n,n}(\mathbf{0}) = \mathcal{W}_{\ell_{n/2},n}(\mathbf{0})$.

Thereafter, using Proposition 2, we obtain $\mathsf{NPB}(f_n) = \mathsf{NPB}(\ell_{n/2})$, that is $\mathsf{NPB}(f_n) = 2^{n/2-1}$ from Proposition 3. Accordingly, using Construction 1 seeded with f_n, we obtain a function g WPB such that $\mathsf{d_H}(f_n, g) = 2^{n/2-1}$.

Since the function $\sigma_{2,n}$ is bent (Property 4) and $\ell_{n/2}$ is affine f_n is also bent (the nonlinearity is an extended affine equivalent criterion), that is $\mathsf{NL}(f_n) = 2^{n-1} - 2^{n/2-1}$. Finally, since the nonlinearity is a distance, the triangular equality gives $\mathsf{NL}(f_n) \leq \mathsf{NL}(g) + \mathsf{d_H}(f_n, g)$ hence $\mathsf{NL}(g) \geq \mathsf{NL}(f_n) - \mathsf{d_H}(f_n, g)$ that is $\mathsf{NL}(g) \geq 2^{n-1} - 2^{n/2}$. □

Remark 2. The proven bound from Proposition 5 is high considering that the nonlinearity of WPB functions is upper bounded by $U_m = 2^{n-1} - 2^{n/2-1} - 2$.

Corollary 1. *Let $m \in \mathbb{N}$, $m \geq 2$ and $n = 2^m$. Let $f_n = \sigma_{2,n} + \ell_{n/2}$ and $C_{k,n} = \mathcal{W}_{f_n,k}(\mathbf{0})/2$. There exist at least*

$$\mathfrak{F}_n = \prod_{k=1}^{n-1} \binom{\frac{1}{2}\binom{n}{k} + |C_{k,n}|}{|C_{k,n}|} \tag{9}$$

WPB functions g such that $\mathsf{NL}(g) \geq 2^{n-1} - 2^{n/2}$.

Proof. Proposition 5 applies Construction 1 with the function $f_n = \sigma_{2,n} + \ell_{n/2}$, where $\ell_{n/2} = \sum_{i=1}^{n/2} x_i$, as input in order to obtain WPB functions with high nonlinearity. We count the number of different functions g that are reachable from f_n, considering the number of different possible support slice by slice. Since for f_n we have that $f_n(0_n) = \ell_{n/2}(0_n) = 0$ and $f_n(1_n) = \ell_{n/2}(1_n) = 0$, Construction 1 always adds 1_n to the support of g (the output WPB function). Then for $k \in [1, n-1]$, recall that $|\mathsf{supp}_k(f_n)| = |\mathsf{E}_{k,n}|/2 - C_{k,n}$ and by construction $|\mathsf{supp}_k(g)| = |\mathsf{E}_{k,n}|/2$. If $C_{k,n} < 0$, we have to subtract a set $\mathsf{S}_{k,n}$ of $|C_{k,n}|$ elements of from $\mathsf{supp}_k(f_n)$. Thus, there are $\binom{|\mathsf{E}_{k,n}|/2-C_{k,n}}{|C_{k,n}|}$ different possible choices for $\mathsf{S}_{k,n}$. If $C_{k,n} > 0$, we have to add $C_{k,n}$ elements to $\mathsf{supp}_k(f_n)$. This corresponds to select a subset $\mathsf{S}_{k,n}$ of $\{\mathsf{E}_{k,n} \setminus \mathsf{supp}_k(f_n)\}$. Hence, we have $\binom{|\mathsf{E}_{k,n}|/2+C_{k,n}}{C_{k,n}}$ possible of choices for $\mathsf{S}_{k,n}$. Therefore, by Construction 1 seeded with f_n we can produce

$$\prod_{k=1}^{n-1} \binom{\frac{1}{2}\binom{n}{k} + |C_{k,n}|}{|C_{k,n}|}$$

different WPB functions with nonlinearity greater than or equal to $2^{n-1} - 2^{n/2}$. □

Applying this construction seeded with f_n we obtain a family of WPB functions with very high nonlinearity for each m. In particular, in the following subsection we discuss the explicit application of the construction as in Proposition 5, and we discuss the results in 8 and 16 variables. Recall that 112 is the maximal nonlinearity obtained experimentally in [MKCL22] through evolutionary algorithm for $n = 8$. Here, we obtain that any function g produced by Construction 1 seeded with $\sigma_{2,8} + \ell_4$ is such that $\mathsf{NL}(g) \geq 112$.

4.1 Concrete examples in 8 and 16 variables

Proposition 5 proves that Construction 1 seeded with the function $f_n = \sigma_{2,n} + \ell_{n/2}$, where $\ell_{n/2} = \sum_{i=1}^{n/2} x_i$, gives WPB functions with high nonlinearity. Implementing this in practice, we were able to construct multiple WPB functions in 8 and 16 with high nonlinearity.

We computed $C_{k,n} = \mathcal{W}_{f_n,k}(\mathbf{0})/2$. For $n = 8$ we have $C_{2,8} = C_{6,8} = 4$, $C_{4,8} = 6$ and for k odd $C_{k,8} = 0$. For $n = 16$ we have $C_{2,16} = C_{14,16} = 8$, $C_{4,16} = C_{12,16} = 28$, $C_{6,16} = C_{10,16} = 56$, $C_{8,16} = 70$ and for k odd $C_{k,16} = 0$. Therefore, from Corollary 1 we know that we can construct more than 2^{41} 8-variable WPB functions with nonlinearity at least 112, and more than 2^{18814} 16-variable WPB functions with nonlinearity at least 32512, as summarized in Table 2.

n	4	8	16
$\mathfrak{F}_n$	6	$> 2^{43}$	$> 2^{1814}$
NL	4	≥ 112	≥ 32512
U_m	4	118	32638

Table 2. Applying Construction 1 seeded with f_n as in Proposition 5 we obtain $\mathfrak{F}_n$ distinct WPB functions. For small values of n we report the size of $\mathfrak{F}_n$, the lower bound on the nonlinearity of these functions from Proposition 5, and the value of the general upper bound from Equation 1.

Explicitly running Construction 1 seeded with f_n we reached nonlinearity up to 116 and 32598 for 8 and 16, respectively. Recall that the theoretical upper bound (1) is $U_3 = 118$ and $U_4 = 32638$, respectively.

Table 3 displays examples of 8-variable WPB functions of this family with nonlinearity value $112, 114$ and 116. Each function is described by providing the points (represented as integers) to join to the support of f_8 for $k < 8$. Similarly, Table 4 contains examples of 16-variables WPB functions of this family with various nonlinearity values, described accordingly.

NL	$\mathsf{S}_{2,8}$	$\mathsf{S}_{4,8}$	$\mathsf{S}_{6,8}$
112	68, 136	90, 105, 204	125, 235
114	40, 129	147, 150, 153	187, 215
116	66, 136	85, 102, 170	123, 215

Table 3. To obtain an 8-variable WPB function g with nonlinearity NL given in the first column, we can set $\mathsf{supp}_k(g) = \mathsf{supp}_k(f_8) \cup \mathsf{S}_{k,8}$ and $g(255) = 1$.

NL	$\mathsf{S}_{2,16}$	$\mathsf{S}_{4,16}$	$\mathsf{S}_{6,16}$	$\mathsf{S}_{8,16}$	$\mathsf{S}_{10,16}$	$\mathsf{S}_{12,16}$	$\mathsf{S}_{14,16}$
32594	2052, 32770, 16416, 514	9256,34884, 8840, 50688, 39424, 36898, 34960, 116, 89, 15, 163,51, 53504, 9346	5702,8998,17801, 8350,31040,10960, 2103,25032,49481, 1861,49812,44545, 12952,10533,16505, 2853,12849,5646, 44552,17177,39712, 32981,6438,9160, 24882,5729,26721	26909,23687,21383, 57902,36398,6331, 14950,14022,44145, 30840,41884,7770, 54452,38580,29081, 50763,30952,45414, 13734,6053,8935, 11827,29739,26195, 20663,30834,27726, 46246,21476,46103, 5215,6042,19341	30073,60660,62196, 44725,30413,30456, 51051,30039,59066, 55786,25335,54963, 64916,55782,55917, 58857,47829,59859, 36813,52907,31356, 58326,46057,3582, 17343,20333,62095,	48093,40863, 55163,31710, 60271,64719, 62460,59381, 65496,61039, 30671,48871, 62439,57069	64383, 32703, 32735, 57341
32596	16388, 272, 16416, 640	16650, 18564, 3077, 17428, 8738, 4481, 9345, 8722, 34945, 4385, 49155, 24588, 165, 16908	49430, 34148, 6482, 12579, 28745, 42784, 5058, 1257, 35341, 35210, 2886, 34438, 25762, 12040, 31008, 37395, 25192, 53300, 14418, 10627, 50340, 20836, 11337, 21168, 41316, 34256, 57425, 2236	29230, 43690, 3960, 13959, 4845, 44818, 44257, 14649, 44182, 7467, 27237, 11162, 45621, 22241, 43417, 27194, 58391, 33501, 25521, 40113, 7051, 55445, 41908, 53713, 21413, 7593, 6068, 14824, 45722, 16823, 879,11956, 38183, 22862,46913	40662, 60075, 47845, 15671, 24181, 28191, 55926, 32593, 8053, 26588, 41663, 42996, 34271, 19679, 8027, 31911, 20410, 33790, 55645, 58842, 14171, 59068, 14139, 52697, 27499, 52188, 55755, 44410	48765, 62439, 57022, 42495, 11775, 30590, 60991, 55271, 65512, 64250, 44975, 28605, 56307, 50943	32763, 57342, 49147, 57215
32598	8256, 2080, 4112, 2049	36912, 5264, 34840, 10264, 49169, 38400, 1632, 3075, 2570, 16800, 16908, 1569, 24612, 12417	29504, 17825, 37413, 18965, 41410, 16613, 5028, 35122, 21656, 61968, 42122, 8000, 24873, 9546, 21541, 10763, 35881, 57372, 45256, 42033, 37524, 19529, 7237, 16446, 17888, 20881, 26817, 49539	14964, 54452, 51612, 22981, 20723, 989, 46868, 50830, 11884, 1518, 5363, 36553, 43729, 39321, 50459, 55401, 37771, 52359, 5965, 8511, 18551, 58538, 14987, 53799, 44090, 10156, 29283, 27057, 58443, 61497, 35782, 44047, 22940, 7540, 19865	43961, 15221, 62179, 43927, 57240, 59741, 61867, 14190, 62511, 44665, 3067, 8107, 61937, 51161, 42937, 31835, 44725, 30435, 14324, 30381, 31964, 56506 54652, 59951, , 61206, 43993, 14310, 58959	32494, 24443, 32381, 62451, 60915, 60381, 44990, 62845, 36351, 32508, 61147, 56309, 32351, 48503	57215, 32751, 63483, 64510

Table 4. To obtain a 16-variable WPB function g with nonlinearity NL as in the first column, we can set $\mathsf{supp}_k(g) = \mathsf{supp}_k(f_{16}) \cup \mathsf{S}_{k,16}$ and $g(65535) = 1$.

5 Distribution of the nonlinearity of WPB functions

In this section we define the notion of distribution of the nonlinearity of WPB functions. Similarly to [GM22a], where the notion of distribution of the WPB functions has been first introduced, we define it as the discrete distribution describing the probability of getting a certain value of nonlinearity by sampling a WPB function uniformly at random:

Definition 12 (Nonlinearity distribution). *Let $m \in \mathbb{N}^*$, $n = 2^m$. The nonlinearity distribution $\mathfrak{N}_n$ is a discrete probability distribution describing the probability of getting a certain value of* NL *by taking a random WPB function, namely for any $x \in \mathbb{N}$*

$$p_{\mathfrak{N}_n}(x) = \frac{|\{f \in \mathcal{WPB}_m \colon \mathsf{NL}(f) = x\}|}{|\mathcal{WPB}_m|}.$$

The support of this distribution is the set of all values that can be realized as nonlinearity of a WPB function. Indeed, $y \in \mathsf{supp}(p_{\mathfrak{N}_n}) = \{a \in \mathbb{N} \colon p_{\mathfrak{N}_n}(a) \neq 0\}$ if and only if there exists $f \in \mathcal{WPB}_m$ such that $\mathsf{NL}(f) = y$. As for the weightwise nonlinearities, this implies that the minimum and maximum nonlinearity of WPB functions are exactly the minimum and maximum of $\mathsf{supp}(p_{\mathfrak{N}_n})$.

The number of 4-variable WPB functions is $|\mathcal{WPB}_2| = 720$. Therefore, applying similar techniques as those in [GM22a], we can retrieve $\mathfrak{N}_4$ in less then 4 seconds on a simple laptop. The distribution is displayed in Figure 1, note that B_2 and U_2 are both reached. However, for larger values of m an exhaustive computation is currently unfeasible, since *e.g.* $|\mathcal{WPB}_3| > 2^{243}$ and $|\mathcal{WPB}_4| > 2^{65452}$. Therefore, when $m > 2$ we can only compute an approximation of $\mathfrak{N}_{2^m}$.

x	2	4
$p_{\mathfrak{N}_n}(x)\%$	16.667	83.333
#	120	600

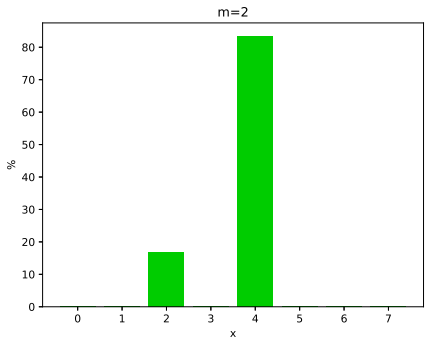

Fig. 1. Distribution $\mathfrak{N}_4$.

5.1 Experimental approximation of $\mathfrak{N}_n$ for $n = 8$ and 16.

To approximate the distribution $\mathfrak{N}_n$ we generate uniformly at random a sample S from $\mathcal{WPB}_m$ and we compute the distribution of the nonlinearity respectively to this sample. In fact, this strategy follows the same principle described in [GM22a] for approximating the distribution of the weightwise nonlinearities. Therefore, we can also apply the same computational techniques based on iterators and parallel computing.

More precisely, let $\mathtt{gen}_\pi(n)$ be a function that returns a random element of $\mathcal{WPB}_m$. Our strategy consists in sampling independently s functions and then computing their nonlinearity, in order to obtain a distribution $\mathfrak{N}'_n$ that is an approximation of $\mathfrak{N}_n$ given by a sample S of size s. Then, we set

$$p_{\mathfrak{N}'_n}(x) = \frac{|\{f \in S \colon \mathsf{NL}(f) = x\}|}{s}.$$

Algorithm 2 illustrates this strategy, denoting in pseudo-code by par-for the fact that the loop is performed in parallel.

Algorithm 2 Approximate nonlinearity distribution of WPB functions

Input: s sample size.
Output: $\mathfrak{N}'_n$.

1: $p = \mathbf{0} \in \mathbb{N}^u$, where u is an upper bound for the max NL.
2: **par-for** $i \in \{1, \ldots, s\}$ **do**
3: $\quad f \leftarrow \mathtt{gen}_\pi(n)$
4: $\quad x \leftarrow \mathsf{NL}(f)$
5: $\quad p_x = p_x + 1$
6: **end for**
7: $\mathfrak{N}'_n = (p_x/s \colon x \in [0, u-1])$
8: **return** $\mathfrak{N}'_n$

Results and remarks. For $n = 8$, we obtained the approximated distribution $\mathfrak{N}'_8$ displayed by Figure 2 and fully summarized by Table 5. $\mathfrak{N}'_8$ with $s > 2^{23}$ samples has maximal and minimal values 112 and 78, respectively. This implies that we can expect functions with nonlinearity outside of this range to be rare. Namely, our experiments show that sampling a random WPB function we can expect its nonlinearity to be almost always close to 104, and almost never larger than 112. However, notice that our approximation provides only a general intuition about the distribution $\mathfrak{N}_8$. Indeed, in Section 4 we prove the existence of 8-variable WPB functions with nonlinearity 8 and others with nonlinearity at least 112. In Section 4.1 we show that actually there exist WPB functions with nonlinearity 112, 114 and 116. More precisely, we provide a family 8-variables WPB functions (of size greater than 2^{43}) having nonlinearity at least 112. For $n = 16$ we obtained an approximated distribution $\mathfrak{N}'_{16}$ by $s > 2^{24}$ samples. $\mathfrak{N}'_{16}$ is summarized by Figure 3 and Table 6. We can expect a WPB function in 16 variables sampled uniformly at random to have nonlinearity close to 32212 and neither smaller than 31886 nor larger than 32300. Again, this is only a general intuition since we prove in Section 4.1 that we can construct more than 2^{1814} functions having nonlinearity at least 32512.

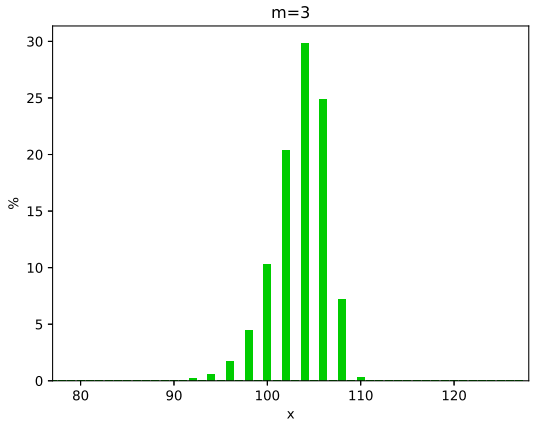

Fig. 2. Approximation of $\mathfrak{N}_8$ via Table 5 data.

x	78	80	82	84	86	88	90	92	94
$p_{\mathfrak{N}_n}(x)\%$	0.000	0.000	0.000	0.001	0.005	0.018	0.063	0.202	0.606
#	2	8	27	115	411	1549	5402	17376	52011
x	96	98	100	102	104	106	108	110	112
$p_{\mathfrak{N}_n}(x)\%$	1.709	4.425	10.308	20.370	29.869	24.897	7.225	0.302	0.000
#	146762	379891	885042	1748852	2564407	2137525	620286	25889	37

Table 5. Approximation of $\mathfrak{N}_8$ via Algorithm 2 with $s = 8585592 > 2^{23}$. See Figure 2.

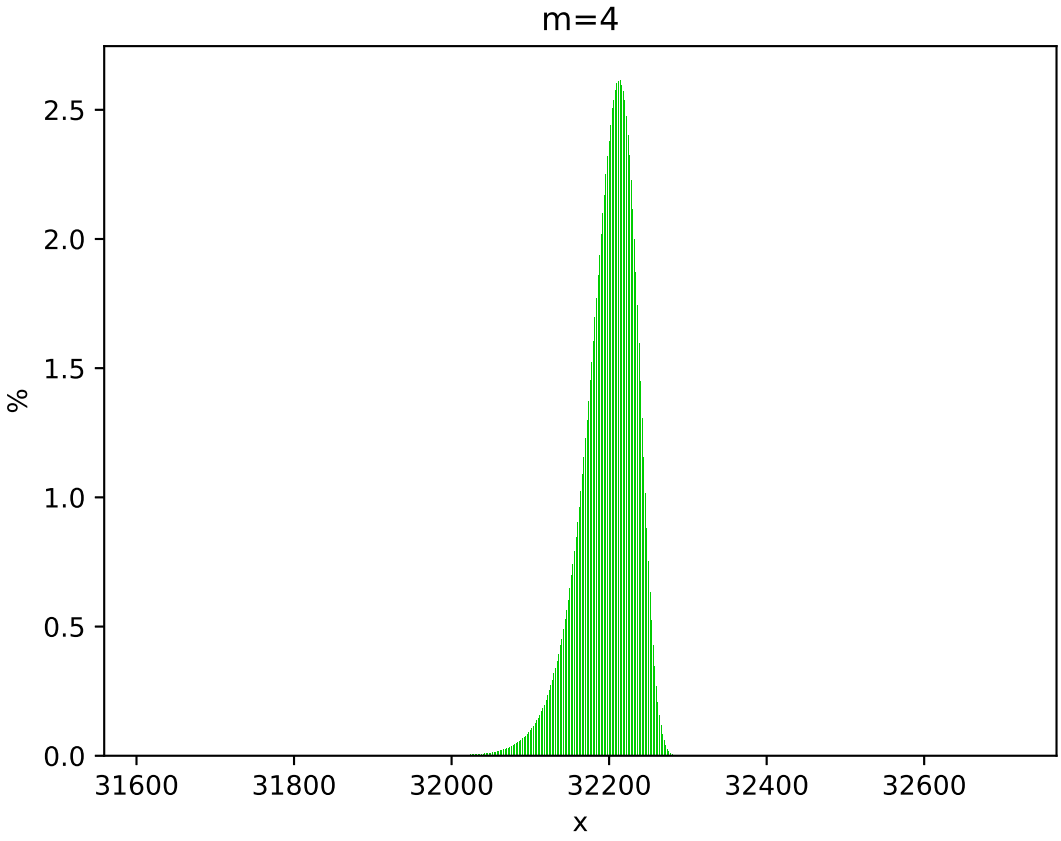

Fig. 3. Approximation of $\mathfrak{N}_{16}$ via Table 6 data.

x	31878	31880	31884	31886	31888	31890	31892	31894	31896	31898	31900	31902
$p_{\mathfrak{N}_n}(x)\%$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
#	1	1	1	2	1	1	2	2	1	1	2	1
x	31904	31906	31908	31910	31912	31916	31918	31920	31922	31924	31926	31928
$p_{\mathfrak{N}_n}(x)\%$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
#	1	1	4	3	1	5	4	11	3	4	8	2
x	31930	31932	31934	31936	31938	31940	31942	31944	31946	31948	31950	31952
$p_{\mathfrak{N}_n}(x)\%$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
#	11	10	7	7	9	11	13	21	13	19	28	19
x	31954	31956	31958	31960	31962	31964	31966	31968	31970	31972	31974	31976
$p_{\mathfrak{N}_n}(x)\%$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
#	25	23	26	35	27	38	50	48	43	60	65	74
x	31978	31980	31982	31984	31986	31988	31990	31992	31994	31996	31998	32000
$p_{\mathfrak{N}_n}(x)\%$	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
#	80	93	95	95	94	147	134	159	169	191	195	240
x	32002	32004	32006	32008	32010	32012	32014	32016	32018	32020	32022	32024
$p_{\mathfrak{N}_n}(x)\%$	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.004
#	254	267	292	325	372	394	441	492	504	541	631	716
x	32026	32028	32030	32032	32034	32036	32038	32040	32042	32044	32046	32048
$p_{\mathfrak{N}_n}(x)\%$	0.004	0.005	0.005	0.005	0.006	0.006	0.007	0.008	0.008	0.009	0.010	0.011
#	777	821	867	981	1080	1129	1272	1410	1500	1626	1804	1943
x	32050	32052	32054	32056	32058	32060	32062	32064	32066	32068	32070	32072
$p_{\mathfrak{N}_n}(x)\%$	0.012	0.013	0.014	0.016	0.018	0.018	0.020	0.022	0.024	0.026	0.029	0.031
#	2088	2357	2587	2814	3238	3342	3615	3958	4413	4735	5196	5576
x	32074	32076	32078	32080	32082	32084	32086	32088	32090	32092	32094	32096
$p_{\mathfrak{N}_n}(x)\%$	0.034	0.036	0.040	0.043	0.047	0.052	0.055	0.060	0.066	0.071	0.077	0.083
#	6087	6570	7185	7760	8446	9462	9971	10862	11963	12786	13960	15004
x	32098	32100	32102	32104	32106	32108	32110	32112	32114	32116	32118	32120
$p_{\mathfrak{N}_n}(x)\%$	0.092	0.098	0.107	0.114	0.124	0.136	0.147	0.160	0.173	0.185	0.198	0.216
#	16643	17679	19410	20710	22377	24605	26541	28948	31245	33485	35878	39032
x	32122	32124	32126	32128	32130	32132	32134	32136	32138	32140	32142	32144
$p_{\mathfrak{N}_n}(x)\%$	0.234	0.253	0.274	0.294	0.318	0.339	0.368	0.395	0.428	0.453	0.490	0.527
#	42349	45765	49650	53182	57629	61402	66557	71506	77511	82064	88690	95420
x	32146	32148	32150	32152	32154	32156	32158	32160	32162	32164	32166	32168
$p_{\mathfrak{N}_n}(x)\%$	0.562	0.602	0.646	0.695	0.739	0.791	0.846	0.902	0.963	1.025	1.089	1.156
#	101851	109070	116977	125949	133882	143285	153179	163369	174408	185596	197313	209364
x	32170	32172	32174	32176	32178	32180	32182	32184	32186	32188	32190	32192
$p_{\mathfrak{N}_n}(x)\%$	1.229	1.297	1.375	1.450	1.523	1.606	1.695	1.773	1.858	1.938	2.018	2.100
#	222611	234952	248929	262600	275734	290871	306909	321097	336415	351029	365403	380364
x	32194	32196	32198	32200	32202	32204	32206	32208	32210	32212	32214	32216
$p_{\mathfrak{N}_n}(x)\%$	2.171	2.251	2.322	2.380	2.442	2.501	2.540	2.576	2.604	2.614	2.608	2.598
#	393218	407735	420543	431073	442322	452937	460029	466583	471623	473496	472316	470558
x	32218	32220	32222	32224	32226	32228	32230	32232	32234	32236	32238	32240
$p_{\mathfrak{N}_n}(x)\%$	2.570	2.539	2.476	2.403	2.326	2.228	2.123	1.999	1.871	1.742	1.592	1.450
#	465477	459756	448388	435231	421306	403534	384544	362026	338889	315447	288252	262553
x	32242	32244	32246	32248	32250	32252	32254	32256	32258	32260	32262	32264
$p_{\mathfrak{N}_n}(x)\%$	1.303	1.155	1.018	0.878	0.751	0.632	0.525	0.430	0.345	0.269	0.207	0.158
#	235999	209170	184350	159016	135944	114468	95066	77860	62407	48738	37488	28564
x	32266	32268	32270	32272	32274	32276	32278	32280	32282	32284	32286	32288
$p_{\mathfrak{N}_n}(x)\%$	0.117	0.083	0.059	0.040	0.027	0.018	0.011	0.007	0.004	0.002	0.001	0.001
#	21264	15050	10746	7251	4951	3222	2054	1248	702	353	222	98
x	32290	32292	32294	32296	32298	32300						
$p_{\mathfrak{N}_n}(x)\%$	0.000	0.000	0.000	0.000	0.000	0.000						
#	61	28	12	6	3	1						

Table 6. Approximation of $\mathfrak{N}_{16}$ via Algorithm 2 with $s = 18110464 > 2^{24}$. See Figure 3.

6 Conclusion

In this article we studied the nonlinearity in the class of WPB functions. First, we discussed two lower bounds on the nonlinearity of a WPB function, introducing also the notion Non Perfect Balancedness (NPB). Then, we presented a new construction of WPB functions with prescribed nonlinearity, and by using this construction we are able to exhibit WPB functions with both low and high non linearity for any n. Finally, we studied the distribution of the nonlinearity of uniform WPB functions.

Up to 16 variables, we analyzed explicitly our construction of WPB functions with almost optimal nonlinearity, and the distribution of the nonlinearity of random functions. We concluded that functions like those we produced have a slim chance to be found by sampling uniformly at random. In Table 7 and Table 8 we summarize the state of the art (including our contributions) about the nonlinearity of WPB functions in 8 and 16 variables, respectively. The symbol * denotes the quantities observed from the approximation of the distributions in our experiments.

Construction	Nonlinearity
Minimum	8
Construction 1 seeded with ℓ_4	8
[TL19]	$[66, 82]$
[CMR17] f_8	88
[GM22b] $g_{6,8}$	96
Average*	103.49
Mode*	104
[MKCL22]	$[110, 112]$
Construction 1seeded with $\sigma_{2,8} + \ell_4$	$[112, 116]$
Upper Bound	118

Table 7. Nonlinearity of 8-variable WPB constructions.

Construction	Nonlinearity
Minimum	128
Construction 1 seeded with ℓ_8	128
[CMR17] f_{16}	29488
[GM22b] $g_{14,16}$	29824
[GM22b] h_{16}	30704
Average*	32199.25
Mode*	32212
Construction 1 seeded with $\sigma_{2,16} + \ell_8$	$[32512, 32598]$
Upper Bound	32638

Table 8. Nonlinearity of 16-variable WPB constructions.

Open questions:

- *WPB functions with higher nonlinearity.* We have seen in Section 5 that there are 4-variable WPB functions reaching U_2. However, in Section 5.1 we did not found WPB functions with nonlinearity reaching U_3, but some attaining 116, which corresponds to the upper bound conjectured by Dobbertin [Dob95] and recently studied in [MMM22]. In 16 variables we did not observe WPB functions reaching this upper bound. A natural question is to determine the maximal nonlinearity of a WPB function, if it is provably lower than U_m or even lower than Dobbertin's bound for m greater than 3.
- *Nonlinearity and addition of symmetric functions.* In the proof of Proposition 5 we have seen that adding a symmetric function (null in 0_n and 1_n) does not modify the NPB of a function. Nevertheless, using Construction 1 with $\ell_{n/2}$ and $\ell_{n/2} + \sigma_{2,n}$ we witnessed that adding $\sigma_{2,n}$ can lead to WPB functions with very high nonlinearity from WPB functions with low nonlinearity. Hence, it would be interesting to determine how evolves the nonlinearity of WPB functions simply by adding symmetric functions.
- *NPB and other cryptographic criteria.* The non perfect balancedness turned out to be crucial for the construction of WPB functions introduced in Section 4. Since its definition is analog to the one of nonlinearity, it engages to study the implications of minimal and maximal NPB on other criteria such as degree, algebraic immunity, and nonlinearity.

Implementation. The concrete results in this paper, in 4, 8 and 16 variables, were computed explicitly via `sagemath` [The17]. We used `BooleanFunction` class from the module `sage.crypto.boolean_function` to encode the functions, and we applied the built-in method nonlinearity based on the Walsh transform. The code of our algorithms, and detailed results of our experiments are available at `https://github.com/agnesegini/WAPB_pub`. Experiments were partially hosted by `https://hpc.uni.lu/` [VBCG14].

Acknowledgments. The two authors were supported by the ERC Advanced Grant no. 787390.

References

BP05. An Braeken and Bart Preneel. On the algebraic immunity of symmetric boolean functions. In *Progress in Cryptology - INDOCRYPT 2005, 6th International Conference on Cryptology in India, Bangalore, India, December 10-12, 2005, Proceedings*, pages 35–48, 2005.

Car04. Claude Carlet. On the degree, nonlinearity, algebraic thickness, and nonnormality of boolean functions, with developments on symmetric functions. *IEEE Trans. Information Theory*, pages 2178–2185, 2004.

Car21. Claude Carlet. *Boolean Functions for Cryptography and Coding Theory*. Cambridge University Press, 2021.

CM21. Claude Carlet and Pierrick Méaux. A complete study of two classes of boolean functions: direct sums of monomials and threshold functions. *IEEE Transactions on Information Theory*, pages 1–1, 2021.

CMR17. Claude Carlet, Pierrick Méaux, and Yann Rotella. Boolean functions with restricted input and their robustness; application to the FLIP cipher. *IACR Trans. Symmetric Cryptol.*, 2017(3), 2017.

CV05. Anne Canteaut and Marion Videau. Symmetric boolean functions. *IEEE Trans. Information Theory*, pages 2791–2811, 2005.

DMS06. Deepak Kumar Dalai, Subhamoy Maitra, and Sumanta Sarkar. Basic theory in construction of boolean functions with maximum possible annihilator immunity. *Designs, Codes and Cryptography*, 2006.

Dob95. Hans Dobbertin. Construction of bent functions and balanced boolean functions with high nonlinearity. In Bart Preneel, editor, *Fast Software Encryption*, pages 61–74, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg.

GM22a. Agnese Gini and Pierrick Méaux. On the weightwise nonlinearity of weightwise perfectly balanced functions. *Discret. Appl. Math.*, 322:320–341, 2022.

GM22b. Agnese Gini and Pierrick Méaux. Weightwise almost perfectly balanced functions: secondary constructions for all n and better weightwise nonlinearities. Cryptology ePrint Archive, Paper 2022/1434, 2022. `https://eprint.iacr.org/2022/1434`.

GS22. Xiaoqi Guo and Sihong Su. Construction of weightwise almost perfectly balanced boolean functions on an arbitrary number of variables. *Discrete Applied Mathematics*, 307:102–114, 2022.

LM19. Jian Liu and Sihem Mesnager. Weightwise perfectly balanced functions with high weightwise nonlinearity profile. *Des. Codes Cryptogr.*, 87(8):1797–1813, 2019.

LS20. Jingjing Li and Sihong Su. Construction of weightwise perfectly balanced boolean functions with high weightwise nonlinearity. *Discret. Appl. Math.*, 279:218–227, 2020.

MCD97. William Millan, Andrew Clark, and Ed Dawson. An effective genetic algorithm for finding highly nonlinear boolean functions. In Yongfei Han, Tatsuaki Okamoto, and Sihan Qing, editors, *Information and Communications Security*, pages 149–158, Berlin, Heidelberg, 1997. Springer Berlin Heidelberg.

Méa21. Pierrick Méaux. On the fast algebraic immunity of threshold functions. *Cryptogr. Commun.*, 13(5):741–762, 2021.

Mes16. Sihem Mesnager. *Bent functions*, volume 1. Springer, 2016.

MJSC16. Pierrick Méaux, Anthony Journault, François-Xavier Standaert, and Claude Carlet. Towards stream ciphers for efficient FHE with low-noise ciphertexts. pages 311–343, 2016.

MKCL22. Sara Mandujano, Juan Carlos Ku Cauich, and Adriana Lara. Studying special operators for the application of evolutionary algorithms in the seek of optimal boolean functions for cryptography. In Obdulia Pichardo Lagunas, Juan Martínez-Miranda, and Bella Martínez Seis, editors, *Advances in Computational Intelligence*, pages 383–396, Cham, 2022. Springer Nature Switzerland.

MMM$^+$18. Subhamoy Maitra, Bimal Mandal, Thor Martinsen, Dibyendu Roy, and Pantelimon Stanica. Tools in analyzing linear approximation for boolean functions related to FLIP. In *Progress in Cryptology - INDOCRYPT 2018*, pages 282–303, 2018.

MMM22. Subhamoy Maitra, Bimal Mandal, and Roy Manmatha. Modifying bent functions to obtain the balanced ones with high nonlinearity. In *Progress in Cryptology - INDOCRYPT*. Springer, 2022.

MPJ$^+$22. Luca Mariot, Stjepan Picek, Domagoj Jakobovic, Marko Djurasevic, and Alberto Leporati. Evolutionary construction of perfectly balanced boolean functions. 2022.

MS78. F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-holland Publishing Company, 2nd edition, 1978.

MS21. Sihem Mesnager and Sihong Su. On constructions of weightwise perfectly balanced boolean functions. *Cryptography and Communications*, 2021.

MSL21. Sihem Mesnager, Sihong Su, and Jingjing Li. On concrete constructions of weightwise perfectly balanced functions with optimal algebraic immunity and high weightwise nonlinearity. *Boolean Functions and Applications*, 2021.

MSLZ22. Sihem Mesnager, Sihong Su, Jingjing Li, and Linya Zhu. Concrete constructions of weightwise perfectly balanced (2-rotation symmetric) functions with optimal algebraic immunity and high weightwise nonlinearity. *Cryptogr. Commun.*, 14(6):1371–1389, 2022.

PCG+16. Stjepan Picek, Claude Carlet, Sylvain Guilley, Julian F. Miller, and Domagoj Jakobovic. Evolutionary algorithms for boolean functions in diverse domains of cryptography. *Evol. Comput.*, 24(4):667–694, dec 2016.

QFLW09. Longjiang Qu, Keqin Feng, Feng Liu, and Lei Wang. Constructing symmetric boolean functions with maximum algebraic immunity. *IEEE Trans. Information Theory*, pages 2406–2412, 2009.

Rot76. Oscar S Rothaus. On "bent" functions. *Journal of Combinatorial Theory, Series A*, 20(3):300–305, 1976.

SM07. Palash Sarkar and Subhamoy Maitra. Balancedness and correlation immunity of symmetric boolean functions. *Discrete Mathematics*, pages 2351 – 2358, 2007.

SZZ93. Jennifer Seberry, Xian-Mo Zhang, and Yuliang Zheng. Nonlinearly balanced boolean functions and their propagation characteristics (extended abstract). In Douglas R. Stinson, editor, *Advances in Cryptology - CRYPTO '93*, volume 773 of *Lecture Notes in Computer Science*, pages 49–60. Springer, 1993.

The17. The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 8.1)*, 2017. `https://www.sagemath.org`.

TL19. Deng Tang and Jian Liu. A family of weightwise (almost) perfectly balanced boolean functions with optimal algebraic immunity. *Cryptogr. Commun.*, 11(6):1185–1197, 2019.

Tok15. Natalia Tokareva. *Bent functions: results and applications to cryptography*. Academic Press, 2015.

VBCG14. Sébastien Varrette, Pascal Bouvry, Hyacinthe Cartiaux, and Fotis Georgatos. Management of an academic HPC cluster: The UL experience. In *2014 International Conference on High Performance Computing & Simulation (HPCS)*, pages 959–967, 2014.

ZS21. Rui Zhang and Sihong Su. A new construction of weightwise perfectly balanced boolean functions. *Advances in Mathematics of Communications*, 0:–, 2021.

ZS22. Linya Zhu and Sihong Su. A systematic method of constructing weightwise almost perfectly balanced boolean functions on an arbitrary number of variables. *Discrete Applied Mathematics*, 314:181–190, 2022.