S-Boxes and Round Functions with Controllable Linearity and Differential Uniformity

Kaisa Nyberg *

Prinz Eugen-Straße 18/6, A-1040 Vienna, Austria

nyberg@ict.tuwien.ac.at

Abstract. In this contribution we consider the stability of linearity and differential uniformity of vector Boolean functions under certain constructions and modifications. These include compositions with affine surjections onto the input space and with affine surjections from the output space, inversions, adding coordinate functions, forming direct sums and restrictions to affine subspaces. As examples we consider some true round function and S-box constructions. More theoretical examples are offered by the bent and almost perfect nonlinear functions. We also include some facts about functions with partially bent components.

1 Introduction

Several methods of constructing S-boxes for an iterated block cipher have been previously presented. The most common methods are based on

- random generation,
- testing against a set of design criteria,
- algebraic constructions having certain good properties,
- or a combination of these.

The round functions typically consist of S-boxes combined in certain ways (e.g. parallel or summing up) and finally the whole cipher is formed by iterating (e.g. DES-like or SPN) certain number of rounds.

At each step of the design of a block cipher algebraic constructions and compositions are used. In this contribution we focus on algebraic properties that are necessary and, in some cases sufficient, to guarantee resistance against the differential and linear cryptanalysis.

For example, in ciphers using small parallel S-boxes the bit permutations between rounds play a crucial role in the security of the cipher (cf. DES and substitution-permutation networks [9]). On the other hand, proven security based

* Sponsored by the Matine Board, Finland.

only on the properties of the round functions can be achieved [18]. In both cases low differential uniformity and low linearity of S-boxes and round functions of iterated block ciphers seem to be necessary conditions and are accepted as useful design criteria of S-boxes and round functions.

Different algebraic methods to combine and modify S-boxes in the construction of a round function have been previously proposed. It is of essential importance to understand how well low differential uniformity and linearity are preserved under different combinations and modifications. We will consider the following:

1. composition with a linear (or affine) surjective mapping onto the input space,
2. composition with a linear (or affine) surjective mapping from the output space,
3. inversion,
4. adding coordinate functions,
5. restriction to a linear (or affine) subspace, and
6. sum of functions with independent inputs

This list is not meant to be exhaustive but represents the most common constructions. For example 4 and 6 are used in the CAST algorithm [2]. We will see that differential uniformity can be controlled under 4, but not linearity. A probabilistic method to overcome these problems was presented in [8].

Modification 2 contains as a special case the chopping of an S-box. It gives a controlled way to modify S-boxes and round functions. It was used in [18] and received special attention in [23]. We will give a simple general treatment of 2. For simplicity, it is assumed that the input and output spaces are linear spaces over $\mathbf{F} = GF(2)$, but the results can be generalized to any finite field.

We conclude that in general, differential uniformity and linearity behave in different ways under the modifications 1–6. As an application of the results on 1–6 we give constructions of round functions of iterated ciphers with proven resistance against differential cryptanalysis, but which can be trivially broken by linear cryptanalysis. Similarily, we show that a cipher can be secure against linear cryptanalysis but easily broken using the differential method.

2 Linearity and Nonlinearity

2.1 Boolean Functions

We denote by $\mathbf{F}$ the finite field $GF(2)$. Let $f : \mathbf{F}^n \to \mathbf{F}$ be a Boolean function. The *nonlinearity* of f is defined as follows [13].

$$\begin{aligned}
&\mathcal{NL}(f) = \min_{A\text{aff.}} \#\{x \in \mathbf{F}^n \mid f(x) \neq A(x)\} \\
&= \min_{L\text{lin.}} \min\{\#\{x \in \mathbf{F}^n \mid f(x) = L(x)\}, 2^n - \#\{x \in \mathbf{F}^n \mid f(x) = L(x)\}\} \\
&= 2^{n-1} - \frac{1}{2} \max_{L\text{lin.}} |\#\{x \in \mathbf{F}^n \mid f(x) = L(x)\} - \#\{x \in \mathbf{F}^n \mid f(x) \neq L(x)\}| \\
&= 2^{n-1} - 2^{n-1} \max_{L\text{lin.}} |c(f, L)|
\end{aligned}$$

where, for $L(x) = L_a(x) = a \cdot x$,

$$\begin{aligned} c(f, L) &= \Pr_X(f(X) = L(X)) - \Pr_X(f(X) \neq L(X)) \\ &= 2(\Pr_X(f(X) = a \cdot X)) - 1/2) \\ &= 2^{-n}\widehat{F}(a) \end{aligned}$$

measures the *correlation* between f and $L = L_a$, and $\widehat{F}$ denotes the Walsh transform of f,

$$\widehat{F}(w) = \sum_{x \in \mathbf{F}^n} (-1)^{f(x)+w \cdot x}, \ w \in \mathbf{F}^n.$$

Various measures of the *linearity* of a Boolean function have been previously used in the literature. In this contribution (see also [17]) we use the following.

Definition 1. The *linearity* of a Boolean function is

$$\mathcal{L}(f) = \max_{L \text{ lin.}} |c(f, L)|.$$

The relationships with the linearity measure Λ_f of Chabaud and Vaudenay [5] and with the linearity measure R_f of Dobbertin [7] are

$$\begin{aligned} \Lambda_f &= 2^{n-1}\mathcal{L}(f), \\ R_f &= 2^n\mathcal{L}(f). \end{aligned}$$

The linearity and nonlinearity are related as follows

$$\mathcal{NL}(f) = 2^{n-1} - 2^{n-1}\mathcal{L}(f). \tag{1}$$

By Parseval's theorem

$$\sum_{L \text{ lin.}} c(f, L)^2 = 1$$

from where it follows that

$$2^{-n/2} \leq \mathcal{L}(f) \leq 1.$$

For n even, the lower bound of linearity is tight and is reached by the *bent functions*. For n odd this lower bound is not reached by any functions, and the general tight lower bound is unknown. For some n, at least for $n = 1, 3, 5$ and 7, the tight lower bound is $2^{-\frac{n-1}{2}}$. For $n = 15$, it was shown in [19] that there exist functions $f : \mathbf{F}^n \to \mathbf{F}$ with $2^{-\frac{n}{2}} < \mathcal{L}(f) = \frac{27}{32}2^{-\frac{n-1}{2}}$. Let $f : \mathbf{F}^n \to \mathbf{F}$ be a function with linearity $\mathcal{L}(f)$. Then the function $g : \mathbf{F}^n \times \mathbf{F}^2 \to \mathbf{F}$, $g(x, y, z) = f(x) + yz$, $x \in \mathbf{F}^n$, $y, z \in \mathbf{F}$, has linearity $\mathcal{L}(g) = 2\mathcal{L}(f)$. Hence for all odd n, $n \geq 15$, there exist Boolean functions f with $2^{-\frac{n}{2}} < \mathcal{L}(f) \leq \frac{27}{32}2^{-\frac{n-1}{2}}$. An important conjecture [7] is that the lower bound is asymptotically tight.

Since bent functions are not balanced, the minimal linearity is not reached by balanced Boolean functions. In fact, the tight lower bound is not known for the balanced Boolean functions. Upper bounds of the minimal linearity of balanced Boolean functions in even dimension can be found in [7] and [21].

2.2 Vector Boolean Functions

From now on we consider a vector Boolean function $f : \mathbf{F}^n \to \mathbf{F}^m$. Let $b \in \mathbf{F}^m$ be a nonzero element, $b = (b_1, \ldots, b_m)$. We denote by $b \cdot f$ the Boolean function, which is the linear combination $b_1 f_1 + \ldots + b_m f_m$ of the coordinate functions $f_1, \ldots, f_m$ of f. Against the usual convention, which is to use the term *component* as a synonyme of *coordinate function*, we will, throughout this paper, call the *nonzero linear combinations $b \cdot f$ of the coordinate functions the components of* f. In [15] the notion of nonlinearity was extended to vector functions as follows. The *nonlinearity of a vector Boolean function* f is $\mathcal{NL}(f) = \min_{b \neq 0} \mathcal{NL}(b \cdot f)$. The following definition is then in full accordance with (1) and extends (1) and the relationship with the measure of Chabaud and Vaudenay to hold also for vector Boolean functions.

Definition 2. The *linearity of a vector Boolean function* f is

$$\mathcal{L}(f) = \max_{b \neq 0} \mathcal{L}(b \cdot f).$$

It follows immediately from the absolute lower bound of linearity of Boolean functions that $\mathcal{L}(f) \geq 2^{-\frac{n}{2}}$. It was proven in [14] that this lower bound is tight if and only if $n \geq 2m$ and n is even. The functions reaching this minimum linearity are called *bent*. In [5] Chabaud and Vaudenay proved the following lower bound of linearity of a vector Boolean function $f : \mathbf{F}^n \to \mathbf{F}^m$.

Theorem 3. [5]

$$\mathcal{L}(f) \geq \frac{1}{2^n}(3 \cdot 2^n - 2 - 2\frac{(2^n - 1)(2^{n-1} - 1)}{2^m - 1})^{1/2} = C(n, m). \tag{2}$$

Observe that $C(n, m)$ is negative if $m = 1$, except for $n = 2$, and

$$C(n, m) < 2^{-\frac{n}{2}}, \text{ if } 1 < m < n - 1$$
$$C(n, m) = 2^{-\frac{n}{2}}, \text{ if } m = n - 1$$
$$C(n, m) = 2^{-\frac{n-1}{2}}, \text{ if } m = n$$
$$C(n, m) > 2^{-\frac{n-1}{2}}, \text{ if } m > n$$

Hence the lower bound $C(n, m)$ cannot be reached if $m < n$ (except for $n = 2$). Neither is it tight for $m > n$ [5]. Indeed, for $\frac{n}{2} < m < n$ and $m > n$ the minimum linearity is unknown.

On the other hand, it is known (see e.g. [15] and [16]) that for $m = n$, functions (even bijective) $f : \mathbf{F}^n \to \mathbf{F}^n$ exist with $\mathcal{L}(f) = 2^{-\frac{n-1}{2}}$. Such functions are called *almost bent* [5]. Almost bent functions exist only, if n is odd, and are characterized by the property that their components have an almost flat correlation spectrum. More precisely, $|c(b \cdot f, L)| = 0$ or $2^{-\frac{n-1}{2}}$ for all $b \in \mathbf{F}^n$, if and only if $f : \mathbf{F}^n \to \mathbf{F}^n$ is almost bent. For example, the power functions $f(x) = x^{2^k+1}$ and $f(x) = x^{2^{2k}-2^k+1}$ in $GF(2^n)$, n odd and $\gcd(n, k) = 1$, have this property [10] and are almost bent.

3 Differential Uniformity

In [16] a function $f : \mathbf{F}^n \to \mathbf{F}^n$ is called *differentially δ-uniform* if

$$\#\{x \in \mathbf{F}^n \mid f(x+a) + f(x) = b\} \leq \delta \text{ for all } a \in \mathbf{F}^n,\ b \in \mathbf{F}^m,\ a \neq 0.$$

Hence the following definition is natural (see also [22]).

Definition 4. *Differential uniformity* $\Delta(f)$ of a function $f : \mathbf{F}^n \to \mathbf{F}^m$ is

$$\Delta(f) = \max_{a \neq 0,\, b} \#\{x \in \mathbf{F}^n \mid f(x+a) + f(x) = b\}.$$

Clearly $\Delta(f) \geq \max\{2, 2^{n-m}\}$. It was shown in [14] that for $m < n$ the minimum differential uniformity 2^{n-m} is reached if and only if $2m \leq n$ and n is even. Such functions are called *perfect nonlinear* and they are the same as the bent functions. For $\frac{n}{2} < m < n$ the minimum differential uniformity is unknown. If $m \geq n$ the minimum differential uniformity is 2. A function which reaches this bound is called *almost perfect nonlinear (APN)* in [18], where examples of such functions are given in the case where m and n are equal and odd.

For $m = n$ even, the minimum differential uniformity is unknown. It was shown in [22] that, for $m = n$ even, there is no APN quadratic permutation. In the next section we generalize this result by repeating the approach of [18] and we show that, for $m = n$ even, there is no APN permutation with partially bent components. Let us mention that no examples of differentially 2-uniform functions are known for $m = n$ even and $n > 2$.

For $m > n$ the minimum differential uniformity is 2, and can be reached by simple modifications of APN functions, as we will see below. Such functions may even have linear components.

We may conclude, that for $m \geq n$ differential uniformity is a weaker notion than linearity. In [5] Chabaud and Vaudenay show that almost bentness implies almost perfect nonlinearity. If $m = n$ odd, the permutation $f : GF(2^n) \to GF(2^n), f(x) = x^{-1}, f(0) = 0$, is differentially 2-uniform without having the minimum nonlinearity. An interesting open question is that of what is the maximum linearity of an APN function when $m = n$.

4 Functions with Partially Bent Components

4.1 Partially Bent Boolean Functions

Functions with quadratic components were studied in [15] and [18] and later in [23]. In this section we adopt the techniques from [5] and generalize the approach of [18] to functions with partially bent components. Such functions have a simple and clear structure and therefore they are useful as illustrative examples of linearity properties.

Definition 5. [4] A Boolean function $f : \mathbf{F}^n \to \mathbf{F}$ is *partially bent*, if there exists a linear subspace U of $\mathbf{F}^n$ such that the restriction of f to U is affine and the restriction of f to any complementary subspace V of U, $V \oplus U = \mathbf{F}^n$, is bent, and f can be represented as a direct sum of the restricted functions, i.e., $f(y+z) = f(y) + f(z)$, for all $z \in U$ and $y \in V$.

The space U is formed by the *linear structures* of f, that is, vectors $\alpha \in \mathbf{F}^n$ such that $f(x+\alpha)+f(x)$ is constant. The dimension ℓ of U is called *linearity dimension* of f [15]. Bent functions exist only in even dimension, hence $n-\ell$ is even.

Let us briefly outline some properties of the autocorrelation function and the Walsh transform of partially bent functions. For more details and other properties we refer to [4].

Let us first consider the autocorrelation function $\widehat{r}$ of a partially bent function f. Let $\alpha \in U$ and $\beta \in V$. Then

$$\begin{aligned}\widehat{r}(\alpha+\beta) &= \sum_{x\in\mathbf{F}^n}(-1)^{f(x+\alpha+\beta)+f(x)} = \sum_{y\in V}\sum_{z\in U}(-1)^{f(y+z+\alpha+\beta)+f(y+z)} \\ &= 2^\ell \sum_{y\in V}(-1)^{f(y+\alpha+\beta)+f(y)} = 2^\ell(-1)^{f(\alpha)+f(0)}\sum_{y\in V}(-1)^{f(y+\beta)+f(y)} \\ &= \begin{cases} 0, \text{if } \beta \neq 0, \\ (-1)^{f(\alpha)+f(0)}2^n, \text{if } \beta = 0. \end{cases}\end{aligned}$$

Hence the autocorrelation function of a partially bent function has the following values

$$\widehat{r}(s) = \begin{cases} 0, \text{if } s \notin U, \\ (-1)^{f(s)+f(0)}2^n, \text{if } s \in U, \end{cases} \tag{3}$$

where U is as in Definition 5. Conversely, if the autocorrelation function of a Boolean function $f : \mathbf{F}^n \to \mathbf{F}$ has only values 0 and $\pm 2^n$, then f is partially bent, which can be seen as follows. The vectors $\alpha \in \mathbf{F}^n$, for which $|\widehat{r}(\alpha)| = 2^n$ are exactly those, for which $f(x+\alpha)+f(x)$ is constant. Clearly, they form a linear subspace U of $\mathbf{F}^n$ and the restriction of f to U is linear. Let V be any complementary subspace of U. Since

$$f(x+z) = f(x)+f(z)+f(0)$$

for all $x \in \mathbf{F}^n$ and $z \in U$, this holds particularly for all $y \in V$ and $z \in U$. It remains to show that the restriction of f to V is bent. Let $\beta \in V$ be not equal to zero. Then $\beta \notin U$ and thus $\widehat{r}(\beta) = 0$. Consequently,

$$\begin{aligned} 0 &= \sum_{x\in\mathbf{F}^n}(-1)^{f(x+\beta)+f(x)} = \sum_{y\in V}\sum_{z\in U}(-1)^{f(y+z+\beta)+f(y+z)} \\ &= \sum_{z\in U}\sum_{y\in V}(-1)^{f(y+\beta)+f(y)} = 2^\ell \sum_{y\in V}(-1)^{f(y+\beta)+f(y)}. \end{aligned}$$

Hence $\sum_{y\in V}(-1)^{f(y+\beta)+f(y)} = 0$, for all $\beta \in V$, $\beta \neq 0$, and therefore the restriction of f to V is bent.

A quadratic Boolean function is partially bent. This follows from the fact that then the difference $f(x+\alpha)+f(x)$ is an affine function of x, for all α, and hence either constant or balanced. Therefore the autocorrelation function of a quadratic function takes only values $\pm 2^n$ and 0.

Let us now calculate the values of the Walsh transform $\widehat{F}$ of a partially bent function $f : \mathbf{F}^n \to \mathbf{F}$. By the Wiener-Khintchin theorem we get

$$\widehat{F}(w)^2 = \sum_{s \in \mathbf{F}^n} \widehat{r}(s)(-1)^{w \cdot s} = \sum_{\alpha \in U} \sum_{\beta \in V} \widehat{r}(\alpha + \beta)(-1)^{w \cdot (\alpha + \beta)}$$
$$= \sum_{\alpha \in U} (-1)^{w \cdot \alpha} = 2^n \sum_{\alpha \in U} (-1)^{f(\alpha) + f(0) + w \cdot \alpha}.$$

Recall that the restriction of f to the ℓ-dimensional linear subspace U is affine. Hence we have

$$\widehat{F}(w)^2 = \begin{cases} 2^{n+\ell}, & \text{if } f(x) + w \cdot x \text{ is constant on } U, \\ 0, & \text{if } f(x) + w \cdot x \text{ is not constant on } U. \end{cases}$$

It follows that the linearity of a partially bent function $f : \mathbf{F}^n \to \mathbf{F}$ is

$$\mathcal{L}(f) = 2^{\frac{\ell - n}{2}},$$

where ℓ is the linearity dimension of f. We also see that f is balanced, i.e. $\widehat{F}(0) = 0$, if and only if the restriction of f is a nonconstant affine function on U, or equivalently, if and only if f has a linear structure $\alpha \in \mathbf{F}^n$ such that $f(x + \alpha) + f(x) = 1$ for all $x \in \mathbf{F}^n$.

4.2 Functions with Partially Bent Components

The purpose of this section is to discuss some basic properties of functions with partially bent components. We also precisize and improve some results from [18], [22] and [23] and simplify their proofs. Examples of functions with partially bent components are offered by the power functions $f(x) = x^{2^k+1}$, $x \in GF(2^n)$, the components of which are quadratic [18].

For a function $f : \mathbf{F}^n \to \mathbf{F}^m$ and vectors $a \in \mathbf{F}^n$, $a \neq 0$, and $b \in \mathbf{F}^m$, we make the following notation.

$$\delta_f(a, b) = \#\{x \in \mathbf{F}^n \mid f(x + a) + f(x) = b\}$$
$$\widehat{r}_f(a, b) = \sum_{x \in \mathbf{F}^n} (-1)^{b \cdot f(x+a) + b \cdot f(x)}.$$

Then (see also [6])

$$\sum_{c \in \mathbf{F}^m} \widehat{r}_f(a, c)(-1)^{c \cdot b} = \sum_{x \in \mathbf{F}^n} \sum_{c \in \mathbf{F}^m} (-1)^{c \cdot f(x+a) + c \cdot f(x) + c \cdot b}$$
$$= 2^m \#\{x \in \mathbf{F}^n \mid f(x + a) + f(x) + b = 0\} = 2^m \delta_f(a, b).$$

Applying the inverse Walsh-Hadamard transform we get

$$\widehat{r}_f(a, c) = \sum_{b \in \mathbf{F}^m} \delta_f(a, b)(-1)^{b \cdot c},$$

and further,

$$\sum_{c \in \mathbf{F}^m} \widehat{r}_f(a,c)^2 = 2^m \sum_{b \in \mathbf{F}^m} \delta_f(a,b)^2. \tag{4}$$

Let us now focus on a special case where $\delta_f(a,b)$ takes at most two values, and the second value (if it exists) is zero. This property is a generalization of perfect nonlinearity and almost perfect nonlinearity, and was introduced and studied in [23].

Lemma 6. *Let $f : \mathbf{F}^n \to \mathbf{F}^m$ be a function and $a \in \mathbf{F}^n$, $a \neq 0$. Let us assume that there is a $\delta > 0$ such that, for all $b \in \mathbf{F}^m$, $\delta_f(a,b) = 0$ or δ. Then*

$$\sum_{b \in \mathbf{F}^m} \widehat{r}_f(a,b)^2 = \delta 2^{n+m}.$$

Proof. Since $\sum_{b \in \mathbf{F}^m} \delta_f(a,b) = 2^n$, the claim follows directly from (4). □

Lemma 7. *Let $f : \mathbf{F}^n \to \mathbf{F}^m$ be a function with partially bent components and $a \neq 0$. Let us assume that there is $\delta > 0$ such that for all $b \in \mathbf{F}^m$, $\delta_f(a,b) = 0$ or δ. Then δ is a power of 2, say $\delta = 2^{n-m+t_a}$, $n - m + t_a \geq 1$, and*

$$\sum_{b \in \mathbf{F}^m} \widehat{r}_f(a,b)^2 = 2^{2n+t_a}. \tag{5}$$

Moreover, a is a linear structure of exactly $2^{t_a} - 1$ components of f.

Proof. The vectors $c \in \mathbf{F}^m$ such that $c \cdot f(x+a) + c \cdot f(x)$ is constant form a linear subspace of $\mathbf{F}^m$. Let t_a be the dimension of this subspace. Then by (3) and Lemma 6

$$\sum_{b \in \mathbf{F}^m} \widehat{r}_f(a,b)^2 = 2^{n-m+t_a} 2^{n+m} = 2^{2n+t_a}.$$

□

Theorem 8. *Let $f : \mathbf{F}^n \to \mathbf{F}^m$ have partially bent components and $\ell_b \geq 0$ be the linearity dimension of the component $b \cdot f$. If there is a $\delta > 0$ such that $\delta_f(a,b) = 0$ or δ, for all $a \in \mathbf{F}^n$, $a \neq 0$, and for all $b \in \mathbf{F}^m$, then there is $t \geq 0$ such that $\delta = 2^{n-m+t}$ and*

$$\sum_{b \neq 0} (2^{\ell_b} - 1) = (2^n - 1)(2^t - 1). \tag{6}$$

Proof. It follows from the assumption and Lemma 7, that (5) holds with $t_a = t$, for all $a \neq 0$. Then by (5)

$$\sum_{b \neq 0} \widehat{r}_f(a,b)^2 = 2^{2n+t} - 2^{2n} = 2^n (2^t - 1),$$

for all $a \neq 0$. By summing this up over $a \neq 0$ we get

$$2^{2n}\sum_{b\neq 0}(2^{\ell_b}-1) = \sum_{b\neq 0}\sum_{a\neq 0}\widehat{r}_f(a,b)^2$$
$$= \sum_{a\neq 0}\sum_{b\neq 0}\widehat{r}_f(a,b)^2 = 2^{2n}(2^t-1)(2^n-1).$$

□

We are using the same notation for n and t as in [23], Section 2.2., and our m corresponds to their s. Hence we can see from (6) that the corresponding unproven formula in [23] is incorrect. Consequently, Theorem 2 of [23] remains unproven. From Theorem 8 we get the following corollary.

Corollary 9. *Assume that n is odd. If there exists a function $f : \mathbf{F}^n \to \mathbf{F}^m$ with $\delta_f(a,b) = 0$ or 2^{n-m+t}, for all $a \neq 0$ and b, then $t \geq 1$, and t and m have the same parity.*

Proof. It follows from (6) that

$$\sum_{b\neq 0}(2^{\ell_b}+1) = (2^n-1)(2^t-1)+2(2^m-1).$$

Since n is odd, all ℓ_b are odd, and the left hand side is divisible by 3 while 2^n-1 is not. Consequently, 3 divides 2^t-1 if and only if 3 divides 2^m-1. □

In the case of odd n and $t = 1$ the equation (6) has exactly one solution, that is, $\ell_b = 1$, for all $b \neq 0$. From this and (4) we get the following result.

Theorem 10. *Let $f : \mathbf{F}^n \to \mathbf{F}^n$ be an almost perfect nonlinear function with partially bent components and n odd. Then each component of f has exactly one nonzero linear structure and the nonzero linear structures of different components are distinct.*

Conversely, if each $a \neq 0$ is a linear structure of exactly one component of a function f with partially bent components, and $m = n$, then by (4)

$$\sum_{b\in\mathbf{F}^m}\delta_f(a,b)^2 = 2^{n+1},$$

for all $a \neq 0$. Since $\sum_{b\in\mathbf{F}^n}\delta_f(a,b) = 2^n$ it follows that $\delta_f(a,b) = 0$ or 2 for all $a \neq 0$ and for all b, that is, f is almost perfect nonlinear.

The case $n - m + t = 1$ was considered in [23] for functions with balanced quadratic components. Based on Theorem 8 we can replace 'quadratic' by 'partially bent'. Particularly, we see that there is no APN permutation with partially bent components in even dimension.

In general, it is not known whether APN functions $f : \mathbf{F}^n \to \mathbf{F}^n$ with partially bent components exist in even dimension, except for $n = 2$, where for example, $f = (f_1, f_2)$,

$$f_1(x_1,x_2) = x_1x_2$$
$$f_2(x_1,x_2) = x_1$$

is clearly an APN function.

For $n = m$ even, and $t = 1$, we get from (6) that the number of bent components is at least $\frac{2}{3}(2^n - 1)$. One solution of (6) is that $\ell_b = 0$ for $\frac{2}{3}(2^n - 1)$ components and $\ell_b = 2$ for $\frac{1}{3}(2^n - 1)$ components. Do such functions exist remains an open problem for $n \geq 4$. The other extreme solution of (6) is that $\ell_b = n$ for one component, which is linear, and $\ell_b = 0$ for all other components. The existence of such functions would imply the existence of a bent function from $\mathbf{F}^n$ to $\mathbf{F}^{n-1}$ which is not possible for $n > 2$.

5 Affine Surjections onto the Input Space

Let $A = L + a : \mathbf{F}^s \to \mathbf{F}^n$ be an affine surjection, where L is a linear surjection. When composed with a function $f : \mathbf{F}^n \to \mathbf{F}^m$ the linearity and differential uniformity are as follows.

Theorem 11.

1. $\mathcal{L}(f \circ A) = \mathcal{L}(f)$
2. $\Delta(f \circ A) = \begin{cases} 2^s, & \text{if } s > n, \text{and} \\ \Delta(f), & \text{if } s = n. \end{cases}$

Proof.
1. It suffices to prove the claim for the components of f. Hence we may assume that $m = 1$. We denote the zero space of L by $\mathrm{Ker}L$. Let V be an n- dimensional linear subspace of $\mathbf{F}^s$ such that $\mathbf{F}^s = V \oplus \mathrm{Ker}L$. Then the restriction of A to V is an affine bijection and every $x \in \mathbf{F}^s$ has a unique representation in the form $x = y + z$, where $y \in V$ and $z \in \mathrm{Ker}L$. Let us denote the Walsh transform of f and $f \circ A$ by $\widehat{F}$ and $\widehat{G}$, respectively. Let $w \in \mathbf{F}^s$ be arbitrary. Then

$$\begin{aligned} \widehat{G}(w) &= \sum_{x \in \mathbf{F}^s} (-1)^{f(Lx+a)+w \cdot x} = \sum_{y \in V} (-1)^{f(Ly+a)+w \cdot y} \sum_{z \in \mathrm{Ker}L} (-1)^{w \cdot z} \\ &= \begin{cases} 2^{s-n} \sum_{y \in V} (-1)^{f(Ly+a)+w \cdot y}, & \text{if } w \cdot z = 0, \text{ for all } z \in \mathrm{Ker}L, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Hence if $\widehat{G}(w) \neq 0$, then $w \cdot z = 0$, for all $z \in \mathrm{Ker}L$. In this case there is a unique $u \in \mathbf{F}^n$ such that $w \cdot x = u \cdot Lx = L^t \cdot x$, for all $x \in \mathbf{F}^s$, where we denote the transpose of L by L^t. This means that w has a unique representation in the form $L^t u$, where $u \in \mathbf{F}^n$, and we have

$$\widehat{G}(w) = 2^{s-n}(-1)^{u \cdot a} \sum_{y \in V} (-1)^{f(Ly+a)+u \cdot (Ly+a)} = 2^{s-n}(-1)^{a \cdot u} \widehat{F}(u).$$

So we have shown that either $\widehat{G}(w) = 0$, or $w = L^t u$ and $|\widehat{G}(w)| = 2^{s-n}|\widehat{F}(u)|$. This proves the first claim.

2. If $s > n$, then there exists an $\alpha \in \mathbf{F}^s$, $\alpha \neq 0$, such that $A(x+\alpha) = A(x)$, for all $x \in \mathbf{F}^s$. Hence $\Delta(f \circ A) = 2^s$. If $s = n$, then for all $\alpha \neq 0$ and for all β

$$\begin{aligned} &\#\{x \in \mathbf{F}^s \mid f(A(x+\alpha)) + f(A(x)) = \beta\} \\ &= \#\{x \in \mathbf{F}^s \mid f(A(x) + A(\alpha) + a) + f(A(x)) = \beta\} \\ &\leq \Delta(f), \end{aligned}$$

and the equality is achieved with a suitable choice of α, since in this case, A is bijective. □

The pitfalls in S-box construction presented in Sections 6.1. and 6.4 of [21] are special cases of the preceding theorem. As far as known to this author affine enlargements of the input space have never been used in the design of S-boxes.

6 Affine Surjections from the Output Space

Let $A = L + a : \mathbf{F}^m \to \mathbf{F}^s$ be an affine surjection, where L is a linear surjection, $a \in \mathbf{F}^s$ and $s \leq m$. When composed with a function $f : \mathbf{F}^n \to \mathbf{F}^m$ the linearity and differential uniformity are as follows.

Theorem 12.

1. $\mathcal{L}(A \circ f) \leq \mathcal{L}(f)$, *with equality if* $s = m$.
2. $\Delta(f) \leq \Delta(A \circ f) \leq 2^{m-s}\Delta(f)$.

Proof.
1. The components of $A \circ f$ form a subset of the components of f plus some constants. More precisely,

$$b \cdot (A \circ f) = (L^t b) \cdot f + b \cdot a.$$

Hence the claim is true and holds with equality, if L is bijective.

2. For $\beta \in \mathbf{F}^s$, let $B = A^{-1}\{\beta + a\}$ denote the preimage set of $\beta + a$ under A. Then

$$\begin{aligned} &\#\{x \in \mathbf{F}^n \mid A(f(x+\alpha)) + A(f(x)) = \beta\} \\ &= \sum_{b \in B} \#\{x \in \mathbf{F}^n \mid f(x+\alpha) + f(x) = b\}. \end{aligned}$$

Since the cardinality of B equals 2^{m-s}, the claim follows. □

6.1 The S-Boxes of MacGuffin

Deletion of output bits of an S-box is a special case of a surjection applied to the output space of a substitution box. By the preceeding theorem the linearity may decrease while the differential uniformity may increase when output bits are deleted. A recent example of this phenomenon is offered by the MacGuffin block

cipher algorithm [3], which makes use of the S-boxes of DES, chopped to the half, i.e., from the original four output bits two are deleted. An analysis of this cipher performed by Rijmen and Preneel [20] shows that linear cryptanalysis of the MacGuffin cipher is about as hard as it is for the DES, but the MacGuffin cipher is slightly weaker against differential cryptanalysis.

However, chopping S-boxes does not always result in a decreased security level. In [18] an example of a DES-like cipher construction is given where the nonlinear substitution function constitutes of one large substitution box, constructed from an almost perfect nonlinear permutation in odd dimension, say 33, by deleting one output bit. If such a cipher has independent round keys and sufficiently many rounds, so that differentials over at least four rounds need to be considered in differential cryptanalysis, then the differential attack is proven to be in average as hard as exhaustive key search [18].

6.2 Chopping of Bent and APN Functions

Let us first consider a bent function $f : \mathbf{F}^n \to \mathbf{F}^m$, $2m \leq n$, n even. It follows immediately from the definition of bent functions that chopping t, $0 \leq t < m$ output coordinates results in increase of differential uniformity by a factor of exactly 2^t, that is, the upperbound of the theorem is reached. However, the chopped function remains perfect nonlinear and bent .

A second example of a function, that preserves linearity and increases differential uniformity by a factor of 2^t if t output bits are deleted, is an APN function with partially bent components in odd dimension.

Theorem 13. *Let $f : \mathbf{F}^n \to \mathbf{F}^n$ be an almost perfect nonlinear function and n odd. Then all components of f are partially bent if and only if $\Delta(A \circ f) = 2^{n-s}\Delta(f)$ for all affine surjections $A : \mathbf{F}^n \to \mathbf{F}^s$ and for all s, $1 \leq s \leq n$.*

Proof. Let us assume first, that all components of f are partially bent. The components of $A \circ f$ form a subset of $2^s - 1$ components of f plus 0 or 1.

Let $\alpha \in \mathbf{F}^n$ be an arbitrary nonzero vector. If α is not a linear structure of any of these components, then $b \cdot ((A \circ f)(x + \alpha)(A \circ f)(x))$ is balanced for all $b \in \mathbf{F}^s$, $b \neq 0$. Therefore (see Appendix)

$$\#\{x \in \mathbf{F}^n \mid (A \circ f)(x + \alpha) + (A \circ f)(x) = \beta\} = 2^{n-s}$$

for all $\beta \in \mathbf{F}^s$. Note that by Theorem 10 there are $2^n - 2^s$ such α. The other $2^s - 1$ vectors are the linear structures of the components of $A \circ f$.

If α is a linear structure of a component, say g_1 of $A \circ f$, then there are $s - 1$ components $g_2, \ldots, g_s$ of $A \circ f$ such that the vector equation

$$(A \circ f)(x + \alpha) + (A \circ f)(x) = \beta$$

is a linear transformation of the system

$$g_i(x + \alpha) + g_i(x) = \gamma_i, \ i = 1, 2, \ldots, s. \tag{7}$$

By Theorem 10 α is not a linear structure of any of $g_2, \ldots, g_s$. Hence $g_i(x + \alpha) + g_i(x)$ is a balanced function of x for all $i = 2, 3, \ldots, s$. Consequently (see Appendix), the number of solutions of (7) is either 2^{n-s+1} or zero. So we have proved that $\Delta(A \circ f) = 2^{n-s+1} = 2^{n-s}\Delta(f)$.

To prove the converse, let us assume that $\Delta(A \circ f) = 2^{n-s}\Delta(f)$ for all affine surjections $A : \mathbf{F}^n \to \mathbf{F}^s$ and for all s, $1 \leq s \leq n$. In fact, we need this only for $s = 1$ and $s = 2$. Applying the assumption in the case where the dimension of the output space of A is one, we get that $\Delta(A \circ f) = 2^n$, which means that every component of f has a linear structure. By Lemma 6 the linear structure is unique for each component.

Let f_0 be an arbitrary component of f. It suffices to show that $f_0(x+\alpha)+f_0(x)$ is a balanced function of x if α is not a linear structure of f_0. Let f_α be the component of f whose linear structure is α. Then by the assumption the system

$$f_0(x + \alpha) + f_0(x) = \beta_0$$
$$f_\alpha(x + \alpha) + f_\alpha(x) = \beta_\alpha$$

has at most 2^{n-1} solutions. Since the second equation holds either never or always, depending only on the value of β_α, it follows that the first equation has always 2^{n-1} solutions, that is, $f_0(x + \alpha) + f_0(x)$ is balanced. Therefore, f_0 is partially bent. □

In the view of this theorem we might extend the definition of APN function to the case $m < n$ saying that a function $f : \mathbf{F}^n \to \mathbf{F}^m$, $m \leq n$, is almost perfect nonlinear, if $\Delta(f) \leq 2^{n-m+1}$.

7 Affine Bijections to the Input Space and from the Output Space

As a corollary of Theorems 11 and 12 we get the following (see also [15] and [16]).

Corollary 14. *Let $f : \mathbf{F}^n \to \mathbf{F}^m$ be a function and let $A : \mathbf{F}^m \to \mathbf{F}^m$ and $B : \mathbf{F}^n \to \mathbf{F}^n$ be linear (or affine) bijections. Then*

1. $\mathcal{L}(A \circ f \circ B) = \mathcal{L}(f)$,
2. $\Delta(A \circ f \circ B) = \Delta(f)$.

8 Inverted Function

The following results were given in [15] and [16] but the proofs were omitted. We take this opportunity to present the simple proofs.

Theorem 15. *If a function $f : \mathbf{F}^n \to \mathbf{F}^n$ is invertible then*

1. $\mathcal{L}(f^{-1}) = \mathcal{L}(f)$,
2. $\Delta(f^{-1}) = \Delta(f)$.

Proof.
1. Let $\widehat{F}_b$ and $\widehat{G}_c$ be the Walsh transforms of $b \cdot f$, $b \neq 0$, and $c \cdot f^{-1}$, $c \neq 0$, respectively. Since f is bijective, we have $\widehat{F}_b(0) = \widehat{G}_c(0) = 0$ for all $b \neq 0$ and $c \neq 0$. Hence it suffices to consider the values of $\widehat{F}_b$ and $\widehat{G}_c$ outside the zero. Let b and c be nonzero vectors in $\mathbf{F}^n$. The claim follows from the following equality.

$$\widehat{F}_b(c) = \sum_{x \in \mathbf{F}^n} (-1)^{b \cdot f(x) + c \cdot x} = \sum_{y \in \mathbf{F}^n} (-1)^{b \cdot y + c \cdot f^{-1}(y)} = \widehat{G}_c(b).$$

2. Since f is a permutation, we have $\#\{x \in \mathbf{F}^n \mid f(x+\alpha)+f(x) = 0\} = 0$, for all $\alpha \in \mathbf{F}^n$, $\alpha \neq 0$. Further, $f(x+\alpha)+f(x) = \beta$ if and only if $f(f^{-1}(y)+\alpha) = y+\beta$, or what is equivalent, $f^{-1}(y+\beta) + f^{-1}(y) = \alpha$, for all $\alpha \neq 0$ and $\beta \neq 0$. This proves the second claim. □

9 Adding Coordinate Functions

Given a function $f : \mathbf{F}^n \to \mathbf{F}^m$ with coordinate functions $f_1, \ldots, f_m$ and a function $g : \mathbf{F}^n \to \mathbf{F}$, we set $\tilde{f} = (f_1, \ldots, f_m, g)$. As a corollary of Theorem 12 we get the following.

Theorem 16.

1. $\mathcal{L}(\tilde{f}) \geq max\ \{\mathcal{L}(f), \mathcal{L}(g)\} \geq \mathcal{L}(f)$
2. $\Delta(f) \geq \Delta(\tilde{f}) \geq \frac{1}{2}\Delta(f)$.

This method has been previously used in the CAST algorithm [2], in which S-boxes of 8 input bits and 32 output bits are constructed by selecting 32 bent Boolean functions in $\mathbf{F}^8$ as coordinate functions. It is exactly as hard for the designer to prove upperbounds to the linearity of such S-box as it is to the cryptanalyst to find the best linear approximation. In [8] the probability that the linearity of such an $m \times n$ S-box is below a given bound is estimated under the assumption *"that the 2^n functions determined from all linear combinations of the n output functions of an S-box may be considered independently in an analysis of their nonlinearities and the probability distribution of the nonlinearity of each function is the same as that of a randomly generated function"*. The estimated probabilities are encouragingly large, but the relevance of the assumption about independence remains an open problem.

10 Direct Sum of Functions

The full substitution function of the CAST algorithm takes 32 input bits and outputs 32 bits, and is constructed by forming the direct sum of four 8×32 S-boxes. The design method of the S-boxes was discussed in the previous section. In this section we discuss the linearity and differential uniformity of direct sums of functions.

Let $f_1 : \mathbf{F}^{n_1} \to \mathbf{F}^m$ and $f_2 : \mathbf{F}^{n_2} \to \mathbf{F}^m$ be two functions. The *direct sum* of f_1 and f_2 is the function $f : \mathbf{F}^{n_1} \times \mathbf{F}^{n_2} \to \mathbf{F}^m$, $f(x, y) = f_1(x) + f_2(y)$. We denote $f = f_1 + f_2$.

Theorem 17.

1. $\mathcal{L}(f_1 + f_2) \leq \mathcal{L}(f_1)\mathcal{L}(f_2)$,
2. $\Delta(f_1 + f_2) \leq \min\{2^{n_2}\Delta(f_1), 2^{n_1}\Delta(f_2), 2^m\Delta(f_1)\Delta(f_2)\}$.

Note that if $m = 1$ then 1. is satisfied with equality and is the same as what sometimes is called the "piling-up lemma".

Proof.

1. Let $b \in \mathbf{F}^m$ be nonzero and let us denote the Walsh transform of $b \cdot (f_1 + f_2)$, $b \cdot f_1$ and $b \cdot f_2$ by $\widehat{F_b}$, $\widehat{G_b}$ and $\widehat{G'_b}$, respectively. It is well known and easy to check that $\widehat{F_b}(u, v) = \widehat{G_b}(u)\widehat{G'_b}(v)$, for all $u \in \mathbf{F}^{n_1}$ and $v \in \mathbf{F}^{n_2}$.

2. Let $\beta \in \mathbf{F}^{n_1}$ and $\gamma \in \mathbf{F}^{n_2}$ be nonzero, and let $\delta \in \mathbf{F}^m$. Then

$$\begin{aligned}
&\#\{(y, z) \mid f_1(y + \beta) + f_1(y) + f_2(z + \gamma) + f_2(z) = \delta\} \\
&= \sum_{z \in \mathbf{F}^{n_2}} \#\{y \in \mathbf{F}^{n_1} \mid f_1(y + \beta) + f_1(y) = f_2(z + \gamma) + f_2(z) + \delta\} \\
&\leq \sum_{z \in \mathbf{F}^{n_2}} \Delta(f_1) = 2^{n_2}\Delta(f_1).
\end{aligned}$$

This gives the first upper bound. The second is obtained from this by changing the roles of f_1 and f_2. We get the third upperbound as follows.

$$\begin{aligned}
&\#\{(y, z) \mid f_1(y + \beta) + f_1(y) + f_2(z + \gamma) + f_2(z) = \delta\} \\
&= \sum_{b \in \mathbf{F}^m} \#\{y \in \mathbf{F}^{n_1} \mid f_1(y + \beta) + f_1(y) = b\} \\
&\qquad \times \#\{z \in \mathbf{F}^{n_2} \mid f_2(z + \gamma) + f_2(z) = \delta + b\} \\
&\leq 2^m \Delta(f_1)\Delta(f_2).
\end{aligned}$$

□

If f_1 and f_2 are bent functions, then their direct sum is a bent function and $\mathcal{L}(f_1 + f_2) = \mathcal{L}(f_1)\mathcal{L}(f_2)$. Morever, all three upperbounds in 2. are reached and are hence equal. Note that in this case m is small compared to n_1 and n_2.

With the CAST algorithm the situation is different. The round function is a direct sum of four S-boxes $f_i : \mathbf{F}^{n_i} \to \mathbf{F}^m$, $i = 1, 2, 3, 4$. Since $m = n_1 + n_2 + n_3 + n_4$, the third upperbound can never be reached. Therefore

$$\Delta(f_1 + f_2 + f_3 + f_4) \leq \min_i 2^{\sum_{j \neq i} n_j} \Delta(f_i).$$

Hence the upperbound only depends of the S-box with least differential uniformity. With the parameters $n_i = 8$ and $m = 32$ this gives the upperbound of 2^{-7} to the probability of the most likely one round differential (characteristic), assuming that the best S-box is differentially 2-uniform.

However, it does not seem likely that the differential uniformity of the round function is as high as indicated by Theorem 17. The reason is the large number of zero entries on the rows of the difference distribution tables. From the proof of the theorem we see that the upperbound is reached if there is a permutation of $\mathbf{F}^m$ originating from a translation (by the vector δ, see above) such that after permuting the columns of the difference distribution table of one function there is a row in this difference distribution table which has nonzero entries exactly at the same locations as some row in the difference distribution table of the second function. This kind of coincidence may be rare. More generally, it might be possible to estimate the expected differential uniformity of the CAST f-function by estimating the expected number of coincidences of locations of nonzero entries.

11 Restrictions to Linear (or Affine) Subspaces

Given a function $f : \mathbf{F}^n \to \mathbf{F}^m$ let $g : \mathbf{F}^s \to \mathbf{F}^m$ be the restriction of f to an s-dimensional affine subspace $a + V$.

Let first $m = 1$. By the linearity $\mathcal{L}(g)$ of the restriction g of a Boolean function f to $a + V$ we mean the maximum value taken over all $w \in \mathbf{F}^n$ of

$$2^{-s}|\#\{x \in a+V \mid f(x) = w \cdot x\} - \#\{x \in a+V \mid f(x) \neq w \cdot x\}|$$
$$= 2^{-s}|\sum_{x \in a+V} (-1)^{f(x)+w\cdot x}| = 2|\Pr_{X \in a+V}(f(X) = w \cdot X) - \frac{1}{2}|.$$

For $m > 1$ we set $\mathcal{L}(g) = \max_{b \neq 0} \mathcal{L}(b \cdot g)$.

Theorem 18.

1. $\mathcal{L}(g) \leq 2^{n-s}\mathcal{L}(f)\}$,
2. $\Delta(g) \leq \Delta(f)$.

Proof.
1. Since the components of g are restrictions of the components of f, it suffices to prove the claim for Boolean functions. Let $\widehat{F}$ be the Walsh transform of f. Then taking the Walsh-Hadamerd transform of $\widehat{F}$ we have

$$(-1)^{f(x)} = 2^{-n} \sum_{t \in \mathbf{F}^n} \widehat{F}(t)(-1)^{t \cdot x}.$$

Using this we get

$$\sum_{x \in a+V} (-1)^{f(x)+w\cdot x} = \sum_{x \in V} (-1)^{f(x+a)+w\cdot x+w\cdot a}$$
$$= 2^{-n} \sum_{x \in V} \sum_{t \in \mathbf{F}^n} \widehat{F}(t)(-1)^{t\cdot(x+a)}(-1)^{w\cdot x+w\cdot a}$$
$$= 2^{-n}(-1)^{w\cdot a} \sum_{t \in \mathbf{F}^n} \widehat{F}(t)(-1)^{t\cdot a} \sum_{x \in V} (-1)^{(t+w)\cdot x}$$

$$= 2^{s-n}(-1)^{w\cdot a} \sum_{t\in w+V^{\perp}} \widehat{F}(t)(-1)^{t\cdot a}$$

where $V^{\perp}$ is the orthogonal subspace of V formed by $v \in \mathbf{F}^n$ such that $v \cdot x = 0$ for all $x \in V$. Then the dimension of $V^{\perp}$ is $n - s$. Consequently,

$$|\sum_{x\in a+V} (-1)^{f(x)+w\cdot x}| \leq 2^{s-n} \sum_{t\in w+V^{\perp}} |\widehat{F}(t)| \leq \max_{t\in\mathbf{F}^n} |\widehat{F}(t)|,$$

which proves the claim.
2. The proof of the second claim follows directly from the definition of differential uniformity. □

Again, bent functions offer examples of functions satisfying the equality. Let us consider the bent function

$$f(x_1, \ldots, x_{2s}) = x_1x_{s+1} + \ldots + x_sx_{2s}.$$

Then $\mathcal{L}(f) = 2^{-s}$, and the linearity of the restricted function to the s-dimensional subspace $x_1 = x_2 = \ldots = x_s = 0$ is equal to $1 = 2^{2s-s}\mathcal{L}(f)$.

Restricted functions occur in DES-like ciphers, where the input data to the round is first expanded, then added to the round key, and then taken as input to the substitution function. Let $f : \mathbf{F}^n \to \mathbf{F}^m$ be the substitution function of a DES- like cipher with nonlinearity $\mathcal{NL}(f)$. Let $E : \mathbf{F}^m \to \mathbf{F}^n$ be a linear expansion mapping. Let k be a fixed round key, and we denote by V_k the affine subspace of $\mathbf{F}^n$ consisting of elements of the form $E(x) + k$, $x \in \mathbf{F}^m$ and by $f|_k$ the restriction of f to V_k. Then

$$|\Pr_X\{b \cdot f(E(X) + k) = a \cdot X\} - \frac{1}{2}| \leq \frac{1}{2}\mathcal{L}(f|_k) \leq \frac{1}{2}2^{n-m}\mathcal{L}(f)$$
$$= \frac{2^{n-1} - \mathcal{NL}(f)}{2^m},$$

to replace an unproven formula in [11], page 152, by a correct one.

12 Examples

Applying the results discussed above let us first show that for all $n < m$ there exists a differentially 2-uniform function $f : \mathbf{F}^n \to \mathbf{F}^m$. If n is odd we can take any APN function from $\mathbf{F}^n$ to $\mathbf{F}^n$ and add sufficiently many new coordinate functions. Then the differential uniformity can only decrease, even if the new coordinates were linear or the same as old components. If n is even, we start with an APN function from $\mathbf{F}^{n+1}$ to $\mathbf{F}^{n+1}$, restrict it to $\mathbf{F}^n$, and then add new coordinate functions if necessary.

The second example is a function $g : \mathbf{F}^n \to \mathbf{F}^n$ such that $\mathcal{L}(g)$ is low but g has a linear structure. Let us start with any function $f : \mathbf{F}^n \to \mathbf{F}^n$ such that $\mathcal{L}(f)$ is small. We denote by $\tilde{f}$ a modification of f which is obtained by deleting one

input coordinate. Then $\mathcal{L}(\tilde{f}) \leq 2\mathcal{L}(f)$. By composing $\tilde{f}$ with a linear projection $L : \mathbf{F}^n \rightarrow \mathbf{F}^{n-1}$ we get a function $g = \tilde{f} \circ L$, such that

$$\mathcal{L}(g) = \mathcal{L}(\tilde{f} \circ L) \leq \mathcal{L}(\tilde{f}) \leq 2\mathcal{L}(f)$$
$$\Delta(g) = 2^n.$$

Since $\mathcal{L}(g)$ is small, a round function of a DES-like cipher can be based on g to guarantee proven resistance against linear attacks [17]. But as it is easy to see, such a cipher has an iterative characteristic with probability 1, that is, a linear structure over the whole cipher, which can be exploited to reduce the complexity of exhaustive key search by a factor of 2.

It is not any harder to give an opposite example, that is, a function $g : \mathbf{F}^n \rightarrow \mathbf{F}^n$ such that $\mathcal{L}(g) = 1$ and $\Delta(g) = 4$. Let us start with any function $f : \mathbf{F}^n \rightarrow \mathbf{F}^n$ such that $\Delta(f)$ is small. We denote by $\tilde{f}$ a modification of f which is obtained by deleting one output coordinate. Then $\Delta(\tilde{f}) \leq 2\Delta(f)$. By replacing the deleted component by the all zero Boolean function, we get a function g such that

$$\mathcal{L}(g) = 1 \text{ and } \Delta(g) \leq 2\Delta(f).$$

Since $\Delta(g)$ is small, a round function of a DES-like cipher can be based on g to guarantee proven resistance against differential attacks [18]. But as it is easy to see, such a cipher has an iterative linear approximation over all rounds of the cipher with probability 1, which can be exploited to determine one bit of the unknown key.

Without going into the details let us mention that it is possible to modify the first example in such a way that the linearity does not increase significantly while the probability of the one-round differential to be iterated is strictly less than one, but is still large enough to give a substantial differential over all but the last round. Then the differential cryptanalysis method can be exploited to search for the last round key exhaustively. Note that if the last round differential holds with probability 1, then there is no way to make distinction between wrong and correct candidates for the last round key.

A similar modification of the second example gives a round function of a DES-like cipher, which is resistant against differential cryptanalysis, but where the last round key can be determined by the linear cryptanalysis method.

References

1. C. Adams and S. E. Tavares, *The structured design of cryptographically good S-boxes*, Journal of Cryptology **3**, 1, 1990, pp. 27-42.
2. C. Adams and S. E. Tavares, *Designing S-boxes for ciphers resistant to differential cryptanalysis*, Proceedings of SPRC'93, Fondazione Ugo Bordoni, 1993.
3. M. Blaze and B. Schneier, *The MacGuffin block cipher algorithm*, these proceedings, pp. 97–110.
4. C. Carlet, *Partially-bent functions*, Advances in Cryptology – CRYPTO'92, Lecture Notes in Computer Science, Springer-Verlag, 1993.
5. F. Chabaud and S. Vaudenay, *Links between differential and linear cryptanalysis*, Proceedings of Eurocrypt'94 (to appear).

6. J. Daemen, *Correlation matrices*, these proceedings, pp. 275–285.
7. H. Dobbertin, *Construction of bent functions and balanced Boolean functions with high nonlinearity*, these proceedings, pp. 61–74.
8. H. M. Heys and S. E. Tavares, *On the security of the CAST encryption algorithm*, to appear in the proceedings of Canadian Conference on Computer and Electrical Engineering, Halifax, September 1994.
9. H. M. Heys and S. E. Tavares, *Substitution-permutation networks resistant to differential and linear cryptanalysis*, 2nd ACM CCCS, Fairfax, Virginia, November 1994.
10. T. Kasami, *Weight enumerators for several classes of the 2nd order binary Reed-Muller codes*, Information and Control 18, 1971.
11. L. Ramkilde Knudsen, *Block ciphers – analysis, design and applications*, Ph.D. thesis, DAIMI PB – 485, November 1994.
12. R. Lidl and H. Niederreiter, "Finite Fields", Encyclopedia of Mathematics and its applications **20**, Addison-Wesley, Reading, Massachusetts, 1983.
13. W. Meier and O. Staffelbach, *Nonlinearity criteria for cryptographic functions*, Advances in Cryptology – EUROCRYPT'89, Lecture Notes in Computer Science, Springer-Verlag (1990), pp. 549-562.
14. K. Nyberg, *Perfect nonlinear S-boxes*, Advances in Cryptology – EUROCRYPT '91, Lecture Notes in Computer Science 547, Springer-Verlag (1991), pp. 378-385.
15. K. Nyberg, *On the construction of highly nonlinear permutations*, Advances in Cryptology – EUROCRYPT '92, Lecture Notes in Computer Science 658, Springer-Verlag (1993), pp. 92-98.
16. K. Nyberg, *Differentially uniform mappings for cryptography*, Advances in Cryptology – EUROCRYPT '93, Lecture Notes in Computer Science 765, Springer-Verlag (1994), pp. 55-64.
17. K. Nyberg, *Linear approximation of block ciphers*, Proceedings of Eurocrypt'94 (to appear).
18. K. Nyberg and L. R. Knudsen, *Provable security against a differential attack*, to appear in J. Crypt. 8, No. 1, 1995 (preliminary version: *Proven security against differential cryptanalysis*, Proceedings of Crypto'92).
19. N. J. Patterson and D. H. Wiedemann, *The covering radius of the* $(2^{15}, 16)$ *Reed-Müller code it at least 16276*, IEEE Trans. on Information Theory **29** (1983), pp. 354–356.
20. V. Rijmen and B. Preneel, *Cryptanalysis of MacGuffin*, these proceedings, pp. 353–358.
21. J. Seberry, X.-M. Zhang and Y. Zheng, *Nonlinearity and probagation characteristics of balanced Boolean functions*, Information and Computation (to appear).
22. J. Seberry, X.-M. Zhang and Y. Zheng, *Nonlinearity characteristics of quadratic substitution boxes*, Proceedings of the Workshop on Selected Areas in Cryptography (SAC '94), May 5-6, 1994, Kingston, Canada. To appear under the title *Relationships among nonlinearity criteria* in the proceedings of EUROCRYPT'94.
23. J. Seberry, X.-M. Zhang and Y. Zheng, *Pitfalls in designing substitution boxes*, Advances in Cryptology – CRYPTO'94, Lecture Notes in Computer Science 839, Springer-Verlag (1994), pp. 383–396.

Appendix: On the Distribution of Values of a Vector Boolean Function

It is well known that a function $f = (f_1, \ldots, f_m) : \mathbf{F}^n \to \mathbf{F}^m$, $1 \leq m \leq n$, over any finite field $\mathbf{F}$ of order q takes all values in $\mathbf{F}^m$ equally many, i.e., q^{n-m} times, if and only if each component $b \cdot f$, $b \neq 0$, takes each value in $\mathbf{F}$ equally many times (see e.g. [12]). In this appendix we give a short proof of this fact in the special case of $\mathbf{F} = GF(2)$. For "concrete" proofs in the case of $\mathbf{F} = GF(2)$ and $m = n$ we refer to [1], and the appendix of the Eurocrypt version of [22].

Let $f : \mathbf{F}^n \to \mathbf{F}^m$ be a function and $\mathbf{F} = GF(2)$. According to [5] we denote by θ_f the characteristic function of f,

$$\theta_f(x, y) = \begin{cases} 1, & \text{if } y = f(x), \\ 0, & \text{otherwise.} \end{cases}$$

Let $b \in \mathbf{F}^m$ and $\widehat{F}_b$ be the Walsh transform of $b \cdot f$. Then

$$\sum_{x,y} \theta_f(x, y)(-1)^{a \cdot x + b \cdot y} = \sum_{x \in \mathbf{F}^n} (-1)^{a \cdot x + b \cdot f(x)} = \widehat{F}_b(a). \tag{8}$$

Applying the inverse Walsh-Hadamard transform with respect to the second variable in $\mathbf{F}^m$, we get

$$\sum_{x \in \mathbf{F}^n} \theta_f(x, y)(-1)^{a \cdot x} = 2^{-m} \sum_{b \in \mathbf{F}^m} \widehat{F}_b(a)(-1)^{b \cdot y}. \tag{9}$$

As an easy application of (8) and (9) we get the proof of the result about uniform distribution of values:

A function $f : \mathbf{F}^n \to \mathbf{F}^m$, $1 \leq m \leq n$, takes each value in $\mathbf{F}^m$ equally many times if and only if each component of f is balanced.

Proof. First, let us observe that by (9) we have for all $y \in \mathbf{F}^m$

$$\#\{x \in \mathbf{F}^n \mid f(x) = y\} = \sum_{x \in \mathbf{F}^n} \theta_f(x, y) = 2^{-m} \sum_{b \in \mathbf{F}^m} \widehat{F}_b(0)(-1)^{b \cdot y}.$$

If each component is balanced, then $\widehat{F}_b(0) = 0$, for all $b \neq 0$, and we get

$$\#\{x \in \mathbf{F}^n \mid f(x) = y\} = 2^{-m} \widehat{F}_0(0) = 2^{n-m},$$

for all $y \in \mathbf{F}^m$.

To prove the converse, let us assume that

$$\#\{x \in \mathbf{F}^n \mid f(x) = y\} = \sum_{x \in \mathbf{F}^n} \theta_f(x, y) = 2^{n-m},$$

for all $y \in \mathbf{F}^m$. Then by (8)

$$\widehat{F}_b(0) = \sum_{y \in \mathbf{F}^m} (-1)^{b \cdot y} \sum_{x \in \mathbf{F}^n} \theta_f(x, y) = \begin{cases} 0, & \text{if } b \neq 0, \\ 2^n, & \text{if } b = 0. \end{cases}$$

□