Chapter 4

S-BOX THEORY

Shannon's concept of product ciphers uses two basic transformations: confusion and diffusion. All modern cryptographic algorithms use in some or other way a collection of S-boxes which provide confusion and P-boxes which spread out the output bits to different S-boxes of the next round. P-boxes have usually a fixed permutation of input and output bits. The strength of product ciphers mainly comes from the "properly" designed S-boxes. Being more precise, having cryptographically "strong" S-boxes, it is relatively easy to design a strong cryptographic algorithm by a careful selection of P-boxes and the number of rounds. "Weak" S-boxes always lead to insecure designs.

Each general cryptographic attack on product ciphers explores some weaknesses in S-boxes. In response, a new S-box criterion is introduced. If the criterion is incorporated into S-boxes, it makes the cryptographic algorithm immune against the attack. For instance, the differential attack caused that a "good" XOR profile was added to the list of S-box criteria.

This chapter outlines the techniques and methods used in designing cryptographically strong S-boxes.

4.1 Boolean Functions

Recall that $\Sigma = \{0, 1\}$. The simplest field which can be defined over Σ is $GF(2) = \langle \Sigma, \oplus, \times \rangle$ with the addition $\oplus$ and multiplication $\times$. $GF(2)$ is called the *binary field*. Clearly, addition is $0 \oplus 0 = 1 \oplus 1 = 0$ and $0 \oplus 1 = 1 \oplus 0 = 1$. Multiplication is defined as $0 \times 0 = 1 \times 0 = 0 \times 1 = 0$ and $1 \times 1 = 1$.

Consider a Boolean function $f : \Sigma^n \rightarrow GF(2)$ which assigns a binary element $y \in \Sigma$ to a vector $x = (x_1, \ldots, x_n) \in \Sigma^n$ of n bits (n-tuple) so $y = f(x)$. For example, the vector space Σ^3 consists of the following vectors:

$$(000), (001), (010), (011), (100), (101), (110), (111)$$

Note that we do not need to use commas to separate components of vectors. For simplicity, we will denote elements (vectors) of Σ^n by their decimal representations used as the subscript so

$$\begin{aligned} \alpha_0 &= (00 \ldots 00) \\ \alpha_1 &= (00 \ldots 01) \\ &\vdots \\ \alpha_{2^n-1} &= (11 \ldots 11) \end{aligned}$$

Let $f : \Sigma^n \rightarrow GF(2)$ be a Boolean function. The binary sequence

$$(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_{2^n-1}))$$

is called the *truth table* of the function f. The sequence with components from $\{1, -1\}$ defined by

$$((-1)^{f(\alpha_0)}, (-1)^{f(\alpha_1)}, \ldots, (-1)^{f(\alpha_{2^n-1})})$$

is called the *sequence* of the function f. A $2^n \times 2^n$ matrix F with entries $f_{i,j} = (-1)^{f(\alpha_i \oplus \alpha_j)}$ is called the *matrix* of the function f.

Let $f(x) = x_1x_2x_3 \oplus x_1x_3 \oplus x_2 \oplus x_3 \oplus 1$ be a function on Σ^3. It is easy to check that

$$f(000) = 1,\ f(001) = 0,\ f(010) = 0,\ f(011) = 1,$$
$$f(100) = 1,\ f(101) = 1,\ f(110) = 0,\ f(111) = 1.$$

So the truth table of f is (10011101) and the sequence of f is $(-1, 1, 1, -1, -1, -1, 1, -1)$ or $(-++---+-)$ where $+$ and $-$ stand for $+1$ and -1, respectively. The matrix of f is

$$F = \begin{bmatrix} - & + & + & - & - & - & + & - \\ + & - & - & + & - & - & - & + \\ + & - & - & + & + & - & - & - \\ - & + & + & - & - & + & - & - \\ - & - & + & - & - & + & + & - \\ - & - & - & + & + & - & - & + \\ + & - & - & - & + & - & - & + \\ - & + & - & - & - & + & + & - \end{bmatrix}.$$

A Boolean function $f : \Sigma^n \to GF(2)$ is said to be *balanced* if its truth table has 2^{n-1} zeros (or ones). For instance, the function $f(x) = x_1x_2 \oplus x_3$; $x \in \Sigma^3$, is balanced since the truth table of f is (01010110) and the function takes the value zero the prescribed 4 times.

A Boolean function $f : \Sigma^n \to GF(2)$ is *affine* if it can be represented in the form

$$f(x_1, \ldots, x_n) = a_0 \oplus a_1x_1 \oplus \cdots \oplus a_nx_n,$$

where $a_i \in \Sigma$ for $i = 0, \ldots, n$. The set of all affine functions over Σ^n is denoted by $\mathcal{A}_n$. An affine function f is called *linear* if $a_0 = 0$. The sequence of an affine (or linear) function is called an affine (or linear) sequence. The function $f(x_1, x_2, x_3) = x_3 \oplus x_1 \oplus 1$ is affine and the function $f(x_1, x_2, x_3) = x_3 \oplus x_1$ is linear.

The *Hamming weight* of a binary vector $\alpha \in \Sigma^n$, denoted by $W(\alpha)$, is the number of ones it contains. For example, $W(010011) = 3$. Given two functions $f, g : \Sigma^n \to GF(2)$, the *Hamming distance* between them is defined as $d(f, g) = W(f(x) \oplus g(x))$, where $W(f(x) \oplus g(x))$ is the weight of the truth table of the function $f(x) \oplus g(x)$. Let $f(x) = x_1x_2$ and $g(x) = x_1 \oplus x_2$ be two Boolean functions. Then

$$d(f, g) = W(f(x) \oplus g(x)) = W(x_1x_2 \oplus x_1 \oplus x_2).$$

As the truth table of the function $f \oplus g = x_1x_2 \oplus x_1 \oplus x_2$ is (0111), the distance $d(f, g) = 3$.

Let $\alpha = (a_1, \ldots, a_n)$ and $\beta = (b_1, \ldots, b_n)$ be two vectors (or sequences), the *scalar product* of α and β, denoted by $\langle \alpha, \beta \rangle$, is defined as the sum of the component-wise multiplications. In particular, when α and β are from Σ^n, $\langle \alpha, \beta \rangle = a_1b_1 \oplus \cdots \oplus a_nb_n$, where the addition and multiplication are over $GF(2)$. If α and β are $(1, -1)$-sequences, the scalar product $\langle \alpha, \beta \rangle = \sum_{i=1}^{n} a_ib_i$, and the addition and multiplication is taken over the reals.

Lemma 4.1 *If $\xi = (a_0, \ldots, a_{2^n-1})$ and $\eta = (b_0, \ldots, b_{2^n-1})$ are the sequences of functions $f_1, f_2 : \Sigma^n \to GF(2)$, respectively, then*

$$\xi * \eta = (a_0b_0, a_1b_1, \ldots, a_{2^n-1}b_{2^n-1})$$

is the sequence of $f_1(x) \oplus f_2(x)$, where $x = (x_1, x_2, \ldots, x_n)$.

Proof: The two sequences are given by $a_i = (-1)^{f_1(\alpha_i)}$ and $b_i = (-1)^{f_2(\alpha_i)}$ for $\alpha_i = 0, \ldots, 2^n - 1$. Then

$$a_i b_i = (-1)^{f_1(\alpha_i)}(-1)^{f_2(\alpha_i)} = (-1)^{f_1(\alpha_i) \oplus f_2(\alpha_i)}.$$

□

Let $f_1(x) = x_1 x_2$ $(x \in \Sigma^2)$ which has its sequence

$$\xi = (-1)^{f_1(0,0)}, (-1)^{f_1(0,1)}, (-1)^{f_1(1,0)}, (-1)^{f_1(1,1)} = (1\ 1\ 1\ -)$$

where $-$ stands for -1. The function $f_2(x) = x_2$ $(x \in \Sigma^2)$ which has the function sequence

$$\eta = (-1)^{f_2(0,0)}, (-1)^{f_2(0,1)}, (-1)^{f_2(1,0)}, (-1)^{f_2(1,1)} = (1\ -1\ -).$$

Now $f_1(x) \oplus f_2(x) = x_1 x_2 \oplus x_2$ has the sequence $(1\ -1\ 1)$ which equals to is $\xi * \eta = (1\ 1\ 1-) * (1\ -1\ -) = (1\ -\ 1\ 1)$.

An $r \times r$ matrix with entries from $\{1, -1\}$ is called a *Hadamard matrix* if $HH^T = rI_r$ where H^t is the transpose of H and I_r is the $r \times r$ identity matrix. It is well known that Hadamard matrices exist when $n = 1, 2$ or n is multiple of 4 [314]. A *Sylvester-Hadamard* or *Walsh-Hadamard matrix* is a $2^n \times 2^n$ matrix H_n which is generated according to the following recursive relation:

$$H_0 = 1,\ H_n = \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix};\ n = 1, 2, \ldots.$$

The way the matrix H_n is constructed from H_{n-1} is written shortly as $H_n = H_1 \otimes H_{n-1}$ where $\otimes$ in means the *Kronecker product*. For instance, let

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \quad B = \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix}$$

then

$$A \otimes B = \begin{bmatrix} a_1 B & a_2 B \\ a_3 B & a_4 B \end{bmatrix} \text{ and } B \otimes A = \begin{bmatrix} b_1 A & b_2 A & b_3 A \\ b_4 A & b_5 A & b_6 A \\ b_7 A & b_8 A & b_9 A \end{bmatrix}$$

An interesting relation between Walsh-Hadamard matrices and the collection of linear functions is described in the next lemma.

Lemma 4.2 *The i-th row (column) of H_n is the sequence of linear function $\varphi_i(x) = \langle \alpha_i, x \rangle$, where $x, \alpha_i \in \Sigma^n$ and α_i is the binary representation of the integer i; $i = 0, 1, \ldots, 2^n - 1$.*

Proof: By induction on n. Let $n = 1$. Note that $H_1 = \begin{bmatrix} + & + \\ + & - \end{bmatrix}$ where $+$ and $-$ stand for 1 and -1, respectively. The first row of H_1 is $\ell_0 = (+\ +)$ which is equal to $\langle 0, x \rangle$. The corresponding function is the constant function $f(x) = 0$. The second row of H_1 is $\ell_1 = (+\ -)$ which is the same as the sequence of $\langle 1, x \rangle$ where $x \in \Sigma$. The corresponding function is $f(x) = x$.

Suppose the lemma is true for $n = 1, 2, \ldots, k-1$. Since $H_k = H_1 \otimes H_{k-1}$, each row of H_n can be written as either (ℓ, ℓ) or $(\ell, -\ell)$ where ℓ is a row in H_{k-1}. From the assumption, ℓ is the sequence of some linear function $\varphi(x)$ where $x = (x_2, \ldots, x_k) \in \Sigma^{k-1}$. Thus (ℓ, ℓ) is the sequence of the function $\phi(y) = \varphi(x)$ where $y = (x_1, \ldots, x_k) \in \Sigma^k$ and $(\ell, -\ell)$ is the sequence of the function $\phi(y) = \varphi(x) \oplus x_1$ where $y = (x_1, \ldots, x_n) \in \Sigma^k$. Thus the lemma is true for k. Since H_k is symmetric, the lemma is also true for columns.

□

The first four Walsh-Hadamard matrices are:

$$H_0 = [1],\ H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

$$H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

$$H_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}.$$

Let $\delta = (i_1, i_2, \ldots, i_p)$ be a constant vector from Σ^p. Then $D_\delta : \Sigma^p \to \Sigma$ is defined as

$$D_\delta(y_1, y_2, \ldots, y_p) = (y_1 \oplus \bar{i_1})(y_2 \oplus \bar{i_2}) \cdots (y_p \oplus \bar{i_p})$$

where $\bar{i_j}$ is the complement of i_j for $j = 1, 2, \ldots, p$. The D-function of δ is useful in obtaining the function representation for the concatenation of binary sequences. Let $f_i : \Sigma^q \to GF(2)$; $i = 0, \ldots, 2^p - 1$, be a collection of 2^p Boolean functions. Also let ξ_i be the sequence of $f_i(x_1, \ldots, x_q)$. Now we create the concatenation ξ of the sequences ξ_i; $i = 0 \ldots, 2^p - 1$ so

$$\xi = (\xi_0, \xi_1, \ldots, \xi_{2^p-1})$$

Obviously, the function which corresponds to ξ, is Boolean function $f : \Sigma^{p+q} \to GF(2)$ and

$$f(y, x) = \bigoplus_{\delta \in \Sigma^p} D_\delta(y) f_{\alpha_\delta}(x) \tag{4.1}$$

where $y = (y_1, \ldots, y_p)$, $x = (x_1, \ldots, x_q)$, and α_δ is the decimal representation of δ. For example, if ξ_1, ξ_2 are the sequences of functions f_1, f_2 ($f_1, f_2 : \Sigma^n \to GF(2)$) then $\xi = (\xi_1, \xi_2)$ is the sequence of the function $g : \Sigma^{n+1} \to GF(2)$ and

$$g(u, x_1, \ldots, x_n) = (1 \oplus u) f_1(x) \oplus u f_2(x).$$

4.2 S-box Design Criteria

There is a set of design criteria which are believed to be essential in the design of cryptographic algorithms. If S-boxes do not satisfy one of the criteria, the cryptographic design based on the S-boxes may be cryptographically weak (or easy to attack). The collection of essential S-box design criteria is:

- Completeness,
- Balance,
- Nonlinearity,
- Propagation criterion, and
- Good XOR profile.

4.2.1 Completeness

The criterion was introduced by Kam and Davida [152]. The criterion is applicable to the whole cryptographic design (or S-P network) rather than a single S-box. Given S-boxes with a fixed structure it is necessary to design a suitable permutation box (P-box) and compute how many rounds are necessary to build up the cross dependencies so any binary output is a complex function of every binary input. The lack of these dependencies enables an opponent to use the "divide and conquer" strategy to analyse the design.

4.2.2 Balance

A Boolean function $f : \Sigma^n \rightarrow GF(2)$ is said to be *balanced* if its truth table has 2^{n-1} zeros (or ones). For instance, $f = x_1x_2 \oplus x_3$, a Boolean function on Σ^3, is balanced since the truth table of f is (01010110) and the function takes the value zero $2^{3-1} = 4$ times. The lack of balance in an S-box causes that each time the S-box is used, it produces outputs with a bias. So some output strings are more probable than other. Even worse as any cryptographic design uses many rounds with the same S-box, the bias tends to accumulate making some output strings less and other more probable. This opens up the design to all sort of attacks which explore a non-uniform output string probability distribution.

Given a balanced function $f : \Sigma^n \rightarrow GF(2)$. What are possible input transformations such that the resulting function preserves the balance.

Lemma 4.3 *Let*

$$g(x) = f(xB \oplus \beta)$$

where B is any $n \times n$ nonsingular matrix and a vector $\beta \in \Sigma^n$. Then g is balanced if and only if f is balanced.

Proof: Note that if B is nonsingular, then for x running through all input values from the set $\{\alpha_0, \ldots, \alpha_{2^n-1}\}$, $y = xB \oplus \beta$ also takes on the same collection of values. Hence if $f(x)$ is balanced so is $g(x) = f(xB \oplus \beta)$ as the output values of g are permuted values of the function f. □

Let $g(x) = f(xB \oplus \beta)$ where $\beta = (1, 1, 1)$ and

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Thus $g(x_1, x_2, x_3) = f(x_1 \oplus x_2 \oplus 1,\ x_1 \oplus x_3 \oplus 1,\ x_2 \oplus 1)$. Clearly, g is also balanced since $g(x_0) = 0$ if and only if $f(x_0B \oplus \beta) = 0$.

Lemma 4.4 *Let $f : \Sigma^n \rightarrow GF(2)$ and $g : \Sigma^m \rightarrow GF(2)$ be Boolean functions. Then the function $h : \Sigma^{n+m} \rightarrow GF(2)$ defined as $h(x, y) = f(x) \oplus g(y)$ is balanced if f is balanced.*

Proof: Observe that $g(\alpha)$ is constant for given α and the truth table of $f(x)$ is zero (one) half the time. Consequently, the truth table of $f(x) \oplus g(y)$ is zero (one) half the time. □

4.2.3 Nonlinearity

The nonlinearity of a Boolean function can be defined as the distance between the function and the set of all affine functions (see Pieprzyk and Finkelstein [236]). More precisely, the *nonlinearity* of a Boolean function $f : \Sigma^n \to GF(2)$ is

$$N_f = \min_{g \in \mathcal{A}_n} d(f, g)$$

where $\mathcal{A}_n$ is the set of all affine functions over Σ^n. Consider the function $f(x_1, x_2) = x_1 x_2$. What is its nonlinearity ? The set $\mathcal{A}_2 = \{0, x_1, x_2, x_1 \oplus x_2, 1, x_1 \oplus 1, x_2 \oplus 1, x_1 \oplus x_2 \oplus 1\}$.

x_1x_2	$f(x) = x_1x_2$	$\ell_1(x) = x_1$	$\ell_2(x) = x_2$	$\ell_3(x) = x_1 \oplus x_2$
00	0	0	0	0
01	0	0	1	1
10	0	1	0	1
11	1	1	1	0

So that $d(f, \ell_1) = d(f, \ell_2) = 1$, $d(f, \ell_3) = 3$. For the missing affine functions the distances are either 1 or 3, so the nonlinearity of f is 1.

Lemma 4.5 *Let $f, g : \Sigma \to GF(2)$ then*

$$d(f, g) = 2^{n-1} - \frac{1}{2}\langle \xi, \eta \rangle$$

where ξ, η are the sequences of f and g, respectively.

Proof: Denote $\xi = (a_0, a_1, \ldots, a_{2^n-1})$ and $\eta = (b_0, b_1, \ldots, b_{2^n-1})$. Let $\rho(+)$ denote the number of positions for which two sequences are the same ($a_j = b_j$). The integer $\rho(-)$ gives the number of positions where the two sequences differ or $a_j \neq b_j$. Hence, $\langle \xi, \eta \rangle = \rho(+) - \rho(-) = 2^n - 2\rho(-)$ and $\rho(-) = 2^{n-1} - \frac{1}{2}\langle \xi, \eta \rangle$. Obviously, $\rho(-) = d(f, g)$. □

The next lemma can be easily verified using the definition of nonlinearity.

Lemma 4.6 *Let ξ be the sequence of a function f on Σ^n. Then the nonlinearity of the function is expressible by*

$$N_f = 2^{n-1} - \frac{1}{2} \max_{i=0,\ldots,2^n-1} \{ | \langle \xi, \ell_i \rangle | \}$$

where ℓ_i is the i-th row of H_n.

Lemma 4.7 *Let f be an arbitrary function on Σ^n. The nonlinearity of f satisfies the following relation*

$$N_f \leq 2^{n-1} - 2^{\frac{1}{2}n-1}.$$

Proof: Let ξ be the sequence of f. Let ℓ_j be the jth row (column) of the Walsh-Hadamard matrix H_n, $j = 0, 1, \ldots, 2^n - 1$. Note that

$$\xi H_n = (\langle \xi, \ell_0 \rangle, \langle \xi, \ell_1 \rangle, \ldots, \langle \xi, \ell_{2^n-1} \rangle).$$

Clearly, $\xi H_n H_n \xi^T = \sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^2$. As $H_n H_n = 2^n I_{2^n}$, $2^n \xi \xi^T = \sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^2$, where I_{2^n} is $2^n \times 2$ identity matrix. The product $\xi \xi^T$ is always equal to 2^n so

$$\sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^2 = 2^{2n}. \qquad (4.$$

The equation (4.2) is called *Parseval's equation* (see [178]). Thus there exist an index j, $0 \le j \le 2^n - 1$, such that $\langle \xi, \ell_j \rangle^2 \ge 2^n$ and equivalently either $\langle \xi, \ell_j \rangle \ge 2^{\frac{1}{2}n}$ or $\langle \xi, \ell_j \rangle \le -2^{\frac{1}{2}n}$.

From Lemma (4.2), ℓ_j is the sequence of some linear function φ_j. For the case $\langle \xi, \ell_j \rangle \ge 2^{\frac{1}{2}n}$, we can use Lemma 4.5 and conclude that $d(f, \varphi_j) \le 2^{n-1} - 2^{\frac{1}{2}n-1}$. For the case $\langle \xi, \ell_j \rangle \le -2^{\frac{1}{2}n}$, we have $\langle \xi, -\ell_j \rangle \ge 2^{\frac{1}{2}n}$. Note that $-\ell_j$ is the sequence of affine function $1 \oplus \varphi_j$. From Lemma 4.5, $d(f, 1 \oplus \varphi_j) \le 2^{n-1} - 2^{\frac{1}{2}n-1}$. So finally we have that $N_f \le 2^{n-1} - 2^{\frac{1}{2}n-1}$. □

The nonlinearity of a Boolean function is invariant under a nonsingular linear transformation.

Lemma 4.8 *Let f be a Boolean function over Σ^n, B be a $n \times n$ nonsingular matrix, and β a constant vector from Σ^n. Then the function $g(x) = f(xB \oplus \beta)$ has the same nonlinearity as the function f so $N_g = N_f$.*

Proof: From the definition of the nonlinearity, there exists an affine function $\varphi(x) \in \mathcal{A}_n$ such that $d(f, \varphi) = N_f$. Consider the function $\psi(x) = \varphi(xB \oplus \beta)$. Obviously $d(g, \psi) = d(f, \varphi)$ and the function ψ is also an affine function i.e. $\psi(x) \in \mathcal{A}_n$. From the definition of nonlinearity, we can deduce that $N_g \le d(g, \psi)$. This proves that $N_g \le N_f$. Since B is nonsingular, the process can be repeated (for B^{-1}) and thus derive that $N_f \le N_g$. □

The notion of nonlinearity can be generalised for a collection of Boolean functions. Let the function $f : \Sigma^n \to \Sigma^m$. The nonlinearity of the function (Nyberg [217]) is

$$N_f = \min_{\alpha \in \Sigma^m, \alpha \neq 0} N_{f_\alpha}$$

where $f_\alpha = \langle \alpha, f \rangle = \alpha_1 f_1 \oplus \cdots \oplus \alpha_m f_m$ is a linear combination of component functions $f = (f_1, \ldots, f_m)$ defined by the vector $\alpha = (\alpha_1, \ldots, \alpha_m)$.

4.2.4 Propagation Criterion

Strict Avalanche Criterion or SAC was introduced by Webster and Tavares [316]. A function $f : \Sigma^n \to GF(2)$ satisfies the SAC if $f(x) \oplus f(x \oplus \alpha)$ is balanced for all α whose weight is 1, i.e. $W(\alpha) = 1$. In other words, the SAC characterises the output when there is a single bit change on the input. *Higher order SAC* is generalisation of the SAC property. Both the SAC and higher order SAC are collectively called propagation criteria ([2],[244]).

We say that f satisfies the *propagation criterion with respect to the vector* α if $f(x) \oplus f(x \oplus \alpha)$ is a balanced function, where $x, \alpha \in \Sigma^n$ and α is a non-zero vector. The function which holds the propagation criterion with respect to all $\alpha \in \Sigma^n$ whose weight is $1 \le W(\alpha) \le k$, is said to satisfy the *propagation criterion of degree k*.

Consider the function $f = x_1 x_2 \oplus x_3$ over Σ^3. Let $\alpha = (1, 1, 0)$. It is easy to check that

$$f(x) \oplus f(x \oplus \alpha) = (x_1 x_2 \oplus x_3) \oplus ((x_1 \oplus 1)(x_2 \oplus 1) \oplus x_3) = x_1 \oplus x_2 \oplus 1$$

is balanced. So f satisfies the propagation criterion with respect to the vector $\alpha = (1, 1, 0)$. Take the following function over Σ^5

$$f(x_1, x_2, x_3, x_4, x_5) = x_1 \oplus x_1 x_5 \oplus x_2 x_4 \oplus x_2 x_5 \oplus x_2 x_4 x_5 \oplus x_3 x_4 x_5.$$

Let the vector $\alpha = (0, 0, 1, 0, 0)$ then the function

$$f(x) \oplus f(x \oplus \alpha) = x_3 x_4 x_5 \oplus (x_3 \oplus 1) x_4 x_5 = x_4 x_5$$

is not balanced. In fact, f does not satisfy the propagation criterion with respect to any vector in the subset

$$\Re = \{(0,0,0,0,0),(0,0,0,0,1),(0,0,0,1,0),(0,0,1,0,0),(0,0,1,1,1)\}.$$

The next theorem shows how a nonsingular linear transformation can be used to obtain a function which satisfies the SAC.

Theorem 4.1 *Let $f : \Sigma^n \to GF(2)$ be a Boolean function and A be an $n \times n$ nonsingular matrix with entries from $GF(2)$. If $f(x) \oplus f(x \oplus \gamma)$ is balanced for each row γ of A, then the function $\psi(x) = f(xA)$ satisfies the SAC.*

For instance, consider the function $f = x_1x_2 \oplus x_3$ which does not satisfy SAC as

$$f(x) \oplus f(x \oplus e_3) = x_1x_2 \oplus x_3 \oplus x_1x_2 \oplus (x_3 \oplus 1) = 1$$

is not balanced, for the vector $e_3 = (001)$. On the other hand,

$$f(x) \oplus f(x \oplus e_1) = x_2,\ f(x) \oplus f(x \oplus e_2) = x_1,\ f(x) \oplus f(x \oplus \gamma) = x_1 \oplus x_2 \oplus 1$$

are balanced for the vectors $e_1 = (100)$, $e_2 = (010)$, $\gamma = (111)$, respectively. Consider the matrix built from these vectors so

$$A = \begin{bmatrix} e_1 \\ e_2 \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

From Theorem (4.1) we conclude that $g(x) = f(xA)$ satisfies the SAC.

Theorem (4.1) can be generalised and used to design a collection of functions each of which satisfying the SAC.

Theorem 4.2 *Let $f_1, \ldots, f_m$ be functions over Σ^n and the set of vectors over Σ^n be*

$$\Re = \{\alpha | f_j(x) \oplus f_j(x \oplus \alpha) \text{ is not balanced for } j,\ 1 \le j \le m\}.$$

If $|\Re| < 2^{n-1}$ then there exists a nonsingular $n \times n$ matrix with entries from $GF(2)$ such that each $\psi_j(x) = f_j(xA)$ satisfies the SAC.

Consider the following three functions $f_1 = x_1 \oplus x_3 \oplus x_2x_3$, $f_2 = x_1 \oplus x_2 \oplus x_1x_2 \oplus x_2x_3$ and $f_3 = x_1x_2 \oplus x_2x_3 \oplus x_1x_3$. The function f_1 does not satisfy the propagation criterion with respect to the vector $(1,0,0)$ only. The function f_2 – to $(1,0,1)$ only and f_3 – to $(1,1,1)$ only. Therefore $\Re = \{(1,0,0), (1,0,1), (1,1,1)\}$ and $|\Re| = 3 < 2^{n-1}$, where $n = 3$. From Theorem 4.2, there exists a nonsingular 3×3 matrix A such that each function $\psi_j(x) = f_j(xA)$ satisfies the SAC. For example, A can be chosen as

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

A Boolean function may not satisfy the propagation criterion. The ultimate failure happens when the function $f(x) \oplus f(x \oplus \alpha)$ is constant. Being more precise, let f be a function over Σ^n. A vector, α, is called a *linear structure* of f if $f(x) \oplus f(x \oplus \alpha)$ is constant. Every function has at least one linear structure – the zero vector. For instance, consider the function $f = x_1x_2 \oplus x_3$ over Σ^3. The vector $\beta = (0,0,1)$ is a linear structure of f as

$$f(x) \oplus f(x \oplus \beta) = (x_1x_2 \oplus x_3) \oplus (x_1x_2 \oplus x_3 \oplus 1) = 1.$$

Needless to say, nonzero linear structures should be avoided in S-boxes as they force the corresponding differences of functions to be constant.

4.2.5 Other Design Criteria

The XOR profile was introduced in Section (3.3.1). The criterion is not very restrictive as the designer of S-boxes needs to take care that XOR profile does not contain entries with "large" numbers. In addition, the XOR profile must be considered in the context of the best round characteristics. It is possible to trade off the largest entries of XOR profile with the number of rounds.

In some circumstances, we may request from a collection of Boolean functions to be linearly nonequivalent [52]. The collection of functions $\{f_1, \ldots, f_m\}$; $f_i : \Sigma^n \to GF(2)$, is linearly nonequivalent if there is no affine transformation for which $f_i(x) = f_j(Ax + \beta)$ where A is an $n \times n$ nonsingular matrix and $\beta \in \Sigma^n$ $(i \neq j)$.

The function $f : \Sigma^n \to GF(2)$ is written in the algebraic normal form if

$$f(x) = a_0 \oplus \sum_{1 \le i \le n} a_i x_i \oplus \sum_{1 \le i < j \le n} a_{ij} x_i x_j \oplus \cdots \oplus a_{12\ldots n} x_1 x_2 \cdots x_n$$

The requirement of short algebraic normal form of a Boolean function becomes essential when the function is too big to be stored as the lookup table. So the function needs to be evaluated "on the fly". Clearly, shorter functions consume less time for their evaluation.

4.3 Bent Functions

In 1976 Rothaus introduced the so-called *bent functions* [258]. Because of their properties, they can be used as building blocks to design Boolean functions with requested properties [1, 162, 226, 324]. Bent functions from $\mathcal{Z}_q^n$ to $\mathcal{Z}_q$ are defined and studied in [163].

A Boolean function f over Σ^n is called bent if

$$2^{-\frac{n}{2}} \sum_{x \in \Sigma^n} (-1)^{f(x) \oplus \langle \beta, x \rangle} = \pm 1$$

for all $\beta \in \Sigma^n$. The expression $f(x) \oplus \langle \beta, x \rangle$ is regarded as a real-valued function.

The following statements are equivalent.

(i) f is bent,

(ii) $\langle \xi, \ell \rangle = \pm 2^{\frac{1}{2}n}$ for any affine sequence ℓ of length 2^n, where ξ is the sequence of f,

(iii) $2^{-\frac{1}{2}n} H_n \xi^T$ is equal to ± 1,

(iv) $f(x) \oplus f(x \oplus \alpha)$ is balanced for any non-zero vector $\alpha \in \Sigma^n$, where $x = (x_1, x_2, \ldots, x_n)$,

(v) the matrix F of the function f is a Hadamard matrix,

(vi) the nonlinearity N_f satisfies $N_f = 2^{n-1} - 2^{\frac{1}{2}n-1}$.

The proof that the statements are equivalent can be found in [1, 274, 324]. Note that the equivalence of (i), (ii), (iii), and (iv) are easy to prove.

As an exercise, we are going to prove that (ii) $\Leftrightarrow$ (vi).

First we prove that (ii) $\Rightarrow$ (vi). Assume that (ii) holds, i.e. $\langle \xi, \ell_j \rangle = \pm 2^{\frac{1}{2}n}$ for each linear sequence ℓ_j of length 2^n and the linear function φ_j corresponds to the linear sequence ℓ_j. Note that $\langle \xi, 1 + \ell_j \rangle = \mp 2^{\frac{1}{2}n}$ for each linear sequence of length 2^n. Note that $1 + \ell_j$ is the sequence of the affine function $1 \oplus \varphi_j$. From Lemma 4.5, for any linear φ_j, either $d(f, \varphi_j) = 2^{n-1} - 2^{\frac{1}{2}n-1}$ or $d(f, 1 \oplus \varphi_j) = 2^{n-1} - 2^{\frac{1}{2}n-1}$. This proves (vi).

Now we prove that (vi) $\Rightarrow$ (ii). This is done by contradiction. Assume that the statement (ii) is false. From Equation (4.2), we can state that there exists a linear sequence ℓ of length 2^n and its linear function φ such that $|\langle \xi, \ell \rangle| > 2^{\frac{1}{2}n}$. Thus either $\langle \xi, \ell \rangle > 2^{\frac{1}{2}n}$ or $\langle \xi, \ell \rangle < -2^{\frac{1}{2}n}$. In the first case, by Lemma 4.5, $d(f, \varphi_j) < 2^{n-1} - 2^{\frac{1}{2}n-1}$ so $N_f < 2^{n-1} - 2^{\frac{1}{2}n-1}$. In the second case, we know that $\langle \xi, -\ell \rangle > 2^{\frac{1}{2}n}$. Note that $-\ell$ is the sequence of the affine function $1 \oplus \varphi$. Using the same argument, we have $d(f, 1 \oplus \varphi_j) < 2^{n-1} - 2^{\frac{1}{2}n-1}$ so $N_f < 2^{n-1} - 2^{\frac{1}{2}n-1}$. This gives the requested contradiction that $N_f \neq 2^{n-1} - 2^{\frac{1}{2}n-1}$ which concludes the proof.

Bent functions have some remarkable properties. Let f be a bent function over Σ^n and ξ be a bent sequence of the function f. Basic properties of bent functions are:

1. n must be even – bent functions exist for even values of n,
2. for $n \neq 2$, the degree of $f \leq \frac{1}{2}n$ – the degree of f written in the algebraic normal form,
3. for any affine function φ, $f \oplus \varphi$ is also bent,
4. $f(xA \oplus \alpha)$ is also bent where A is any nonsingular matrix of order n, and α is any vector in Σ^n,
5. f takes the value zero $2^{n-1} \pm 2^{\frac{1}{2}n-1}$ times,
6. $2^{-\frac{1}{2}n} H_n \xi^T$ is also a bent sequence.

We now verify some of the properties for the bent function $f(x) = x_1 x_2$ over Σ^2.

- The truth table of f has to contain $2^1 \pm 2^0$ ones (or zeros). As $f(0,0) = 0$, $f(0,1) = 0$, $f(1,0) = 0, f(1,1) = 1$ the truth table is (0001) so the weight of it is 1.
- The 4×4 Sylvester-Hadamard matrix is

$$H_2 = \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix}.$$

  The sequence of $f = x_1 x_2$ is $\xi = (+\ +\ +\ -)$. It is easy to compute that $\langle \xi, \ell_1 \rangle = 2$, $\langle \xi, \ell_2 \rangle = 2$, $\langle \xi, \ell_3 \rangle = 2$, and $\langle \xi, \ell_4 \rangle = -2$. This property is consistent with the statement (ii).

- The matrix of f is

$$F = \begin{bmatrix} + & + & + & - \\ + & + & - & + \\ + & - & + & + \\ - & + & + & + \end{bmatrix},$$

  which is a Hadamard matrix as $FF^T = 4I_4$.

- According to Statement (iv) $f(x) \oplus f(x \oplus \alpha)$ has to be balanced for all nonzero $\alpha \in \Sigma^2$. Indeed, $f(x) \oplus f(x \oplus \alpha) = x_1 x_2 \oplus (x_1 \oplus a_1)(x_2 \oplus a_2) = a_1 x_2 \oplus a_2 x_1 \oplus a_1 a_2$ is an affine function thus 0-1 balanced.

Consider another bent function $f = x_1 x_2 \oplus x_3 x_4$ over Σ^n. The truth table of f is

$$0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 1,\ 0.$$

The function f takes on the value zero $2^{4-1} + 2^{\frac{1}{2}4-1} = 8 + 2 = 10$ times. The function is not balanced.

4.4 Propagation and Nonlinearity

There is an intrinsic relation between propagation properties and the nonlinearity of Boolean functions. For instance, bent functions satisfy propagation criterion with respect to all nonzero vectors. Now we are going to investigate the relation between propagation and nonlinearity for arbitrary Boolean functions.

Let f be a function over Σ^n and $\xi(\alpha)$ be the sequence of the function $f(x \oplus \alpha)$. Using our notation, it is obvious that $\xi(0) * \xi(\alpha)$ is the sequence of $f(x) \oplus f(x \oplus \alpha)$. The *autocorrelation* of f with a shift α is defined as

$$\Delta(\alpha) = \langle \xi(0), \xi(\alpha) \rangle.$$

Lemma 4.9 *Let f be a function over Σ^n. Then the Hamming weight of $f(x) \oplus f(x \oplus \alpha)$ is equal to $2^{n-1} - \frac{1}{2}\Delta(\alpha)$.*

Proof: Let e_+ (e_-) denote the number of ones (minus ones) in the sequence of $\xi(0) * \xi(\alpha)$. Thus $e_+ - e_- = \Delta(\alpha)$ and $(2^n - e_-) - e_- = \Delta(\alpha)$ so $e_- = 2^{n-1} - \frac{1}{2}\Delta(\alpha)$. Note that e_- is also the number of ones in the truth table of $f(x) \oplus f(x \oplus \alpha)$. Thus the lemma holds. □

The following corollary is a simple conclusion from Lemma (4.9).

Corollary 4.1 *$\Delta(\alpha) = 0$ if and only if $f(x) \oplus f(x \oplus \alpha)$ is balanced, i.e. f satisfies the propagation criterion with respect to α.*

Note that if $|\Delta(\alpha)| = 2^n$ then $f(x) \oplus f(x \oplus \alpha)$ is constant and then α is a linear structure (see [217]). In practice, for most Boolean functions, the propagation criterion with respect to arbitrary α is not satisfied and also α is not a linear structure. For some cases, $\Delta(\alpha) \neq 0$ and is relatively small so $f(x) \oplus f(x \oplus \alpha)$ is almost balanced, and the function f has "good" propagation properties. To measure the global propagation property of a function f with respect to all vectors in Σ^n, we can use the number

$$\sum_{\alpha \in \Sigma^n} \Delta^2(\alpha).$$

Ideally, we expect the number to be as small as possible. In fact, it is smallest for bent functions and largest for affine functions.

Let F be the matrix of $f : \Sigma^n \to GF(2)$, ξ be the sequence of f. It is easy to verify that the first row of FF^T is

$$(\Delta(\alpha_0), \Delta(\alpha_1), \cdots, \Delta(\alpha_{2^n-1})).$$

Now consider the Fourier transform of the function f written in the form $2^{-n} H_n F H_n$. According to the result by McFarland (see Theorem 3.3 of [93]), the matrix F can be represented as

$$F = 2^{-n} H_n diag(\langle \xi, \ell_0 \rangle, \cdots, \langle \xi, \ell_{2^n-1} \rangle) H_n \tag{4.3}$$

where ℓ_i is the i-th row of a Sylvester-Hadamard matrix H_n and $diag(a_0, \cdots, a_{2^n-1})$ is a $2^n \times 2^n$ matrix with all zero entries except for the diagonal whose entries are $(a_0, \cdots, a_{2^n-1})$. Using Equation (4.3), the matrix FF^T takes on the form

$$FF^T = 2^{-n} H_n diag(\langle \xi, \ell_0 \rangle^2, \cdots, \langle \xi, \ell_{2^n-1} \rangle^2) H_n$$

Note that f and H_n are symmetric so $F = F^t$ and $H_n = H_n^t$. The first row of FF^t is

$$2^{-n}(\langle \xi^*, \ell_0 \rangle, \cdots, \langle \xi^*, \ell_{2^n-1} \rangle) = 2^{-n} \xi^* H_n$$

where $\xi^* = (\langle \xi, \ell_0 \rangle^2, \cdots, \langle \xi, \ell_{2^n-1} \rangle^2)$. Thus

$$(\Delta(\alpha_0), \Delta(\alpha_1), \cdots, \Delta(\alpha_{2^n-1})) = 2^{-n}(\langle \xi, \ell_0 \rangle^2, \cdots, \langle \xi, \ell_{2^n-1} \rangle^2) H_n.$$

So the following theorem has been proved.

Theorem 4.3 *Let f be a function over Σ^n. Then*

$$(\Delta(\alpha_0), \Delta(\alpha_1), \cdots, \Delta(\alpha_{2^n-1})) H_n = (\langle \xi, \ell_0 \rangle^2, \cdots, \langle \xi, \ell_{2^n-1} \rangle^2).$$

As $\langle \xi, \ell_i \rangle$ expresses the distance between the function f and the linear function which corresponds to the sequence ℓ_i, Theorem (4.3) characterises the relation between the nonlinearity and the propagation. Let us investigate the relation in more details. First denote $\eta = (\Delta(\alpha_0), \Delta(\alpha_1), \cdots, \Delta(\alpha_{2^n-1}))$. The expression $\langle \xi^*, \xi^* \rangle = \langle \eta H_n, \eta H_n \rangle = \eta H_n H_n^T \eta^T = 2^n \langle \eta, \eta \rangle$. As $\langle \xi^*, \xi^* \rangle = \sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^4$, we have shown that the following corollary is true.

Corollary 4.2 *Let f be a function over Σ^n. Then*

$$\sum_{\alpha \in \Sigma^n} \Delta^2(\alpha) = 2^{-n} \sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^4.$$

Corollary (4.2) gives an insight into the relation between propagation properties expressed by $\Delta(\alpha)$ and the nonlinearity characterised by distances of f to the set of linear functions. Clearly, the larger the nonlinearity of f the better the propagation of the function. It is convenient to describe the nonlinearity and propagation of the function f by the parameter

$$\sigma(f) = \sum_{\delta \in \Sigma^n} \Delta^2(\alpha) = 2^{-n} \sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^4. \tag{4.4}$$

the parameter $\sigma(f)$ is called the *global propagation* of the function f. It would be interesting to know how it behaves depending on the function f. The next theorem gives the answer.

Theorem 4.4 *Let f be a function over Σ^n. Then*

(i) $2^{2n} \leq \sigma(f) \leq 2^{3n}$,

(ii) $\sigma(f) = 2^{2n}$ *if and only if f is a bent function,*

(iii) $\sigma(f) = 2^{3n}$ *if and only if f is an affine function.*

Proof: Statement (i). By the definition, we have

$$\sigma(f) = 2^{-n} \sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^4 \leq 2^{-n} (\sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^2)^2.$$

From Equation (4.2), we have

$$\sum_{j=0}^{2^n-1} \langle \xi, \ell_j \rangle^2 = 2^{2n}.$$

Thus $\sigma(f) \leq 2^{-n} 2^{4n} = 2^{3n}$.

Statement (ii). Note that always $\Delta(0) = 2^n$. So $\sigma(f) = \sum_{\alpha \in \Sigma^n} \Delta^2(\alpha) = \Delta^2(0) = 2^{2n}$ happens if and only if $\Delta(\alpha) = 0$ for any $\alpha \neq 0$. This means that f is bent.

Statement (iii). Denote $y_j = \langle \xi, \ell_j \rangle^2$. By Parseval's equation, $\sum_{j=0}^{2^n-1} y_j = 2^n$. The following statements are equivalent: $\sigma(f) = 2^{3n} \iff 2^{-n} \sum_{j=0}^{2^n-1} y_j^2 = 2^{3n} \iff \sum_{j=0}^{2^n-1} y_j^2 = 2^{4n} \iff \sum_{j=0}^{2^n-1} y_j^2 = (\sum_{j=0}^{2^n-1} y_j)^2 \iff y_i y_j = 0$ if $j \neq i \iff$ there exists a j_0 such that $y_{j_0} = 2^{2n}$ and $y_j = 0$ if $j \neq j_0 \iff$ there exists a j_0 such that $\langle \xi, \ell_{j_0} \rangle = \pm 2^n$ and $\langle \xi, \ell_j \rangle = 0$ if $j \neq j_0 \iff$ there exists a j_0 such that $\xi = \pm \ell_{j_0}$ i.e. f is an affine function. □

4.5 Constructions of Balanced Functions

Bent functions have the largest nonlinearity and good propagation properties but are not balanced so they cannot be used in most cryptographic designs. We study two methods of the construction of balanced functions. The first method concatenates bent functions. The second one applies linear functions. Readers interested in details are referred to [274, 275, 276, 277].

4.5.1 Concatenating Bent Functions

There are two case. The first one when we have a bent function over Σ^{2k} and we would like to construct balanced functions over Σ^{2k+1}. The second case is when we want to construct balanced functions over Σ^{2k+2} having a bent function over Σ^{2k}. Our considerations start from the first case.

Let $f : \Sigma^{2k} \to GF(2)$ be a bent function and g be a function over Σ^{2k+1} defined by

$$g(x_1, x_2, \ldots, x_{2k+1}) = x_1 \oplus f(x_2, \ldots, x_{2k+1}).$$

Incidentally, this construction is embedded in cubing permutations over $GF(2^{2k+1})$ (see [235]). The function g is balanced as its truth table is the concatenation of the truth tables of the original function f and its negation, i.e. the function $f \oplus 1$. The function g satisfies the propagation criterion with respect to all non-zero vectors $\alpha \in \Sigma^{2k+1}$ and different from $(1, 0, \ldots, 0)$. This happens as $g(x) \oplus g(x \oplus \alpha)$ is balanced for all $\alpha \notin \{(0, \ldots, 0), (1, 0, \ldots, 0)\}$ (or $\Delta(\alpha) = 0$). If $\alpha = (1, 0, \ldots, 0) = \alpha_1$, then $g(x) \oplus g(x \oplus \alpha_1) = 1$ for all $x \in \Sigma^{2k+1}$ and $\Delta(\alpha_1) = -2^{2k+1}$. The vector α_1 is a nonzero linear structure of g. The global propagation $\sigma(g)$ can be calculated and

$$\sigma(g) = \sum_{\alpha \in \Sigma^{2k+1}} \Delta^2(\alpha) = \Delta^2(0) + \Delta^2(\alpha_1) = 2 \cdot 2^{4k+2} = 2^{4k+3}.$$

The lower bound of $\sigma(h)$, where h is a function on Σ^{2k+1}, is 2^{4k+2}. This bound is attained by bent functions only. But bent functions exist in even dimension vector spaces only.

Denote $g^*(x) = g(xA)$, where A is a nonsingular $(2k+1) \times (2k+1)$ matrix with entries from $GF(2)$. The function g^* is a balanced function on Σ^{2k+1}. Note that $\sigma(g)$ is invariant under any nondegenerate linear transformation on the variables. Thus $\sigma^*(g^*) = 2^{4k+3}$. Clearly, the nonlinearity and the number of vectors for which the propagation criterion is satisfied, is the same for g^* and g. Unfortunately, g has a linear structure although it satisfies the propagation criterion with respect to other nonzero vectors.

Let f be a bent function over Σ^{2k-2} and g be a function over Σ^{2k} defined by

$$g(x_1, x_2, \ldots, x_{2k}) = x_1 \oplus x_2 \oplus f(x_3, \ldots, x_{2k}).$$

For any nonzero vector $\alpha \in \Sigma^{2k}$, consider $g(x) \oplus g(x \oplus \alpha)$. Denote $\alpha_1 = (1, 0, \ldots, 0)$, $\alpha_2 = (0, 1, \ldots, 0)$, $\alpha_3 = (1, 1, \ldots, 0)$. First assume that $\alpha \neq \alpha_1, \alpha_2, \alpha_3$. From the definition of the function g, it is easy to conclude that $g(x) \oplus g(x \oplus \alpha)$ is balanced and $\Delta(\alpha) = 0$. On the other hand, suppose that $\alpha = \alpha_j$, $j = 1, 2, 3$. From the definition of g, we have that $g(x) \oplus g(x \oplus \alpha_j) = 1; j = 1, 2$, for all $x \in \Sigma^{2k+1}$. Also $\Delta(\alpha_j) = -2^{2k}$ for $j = 1, 2$. For α_3, $g(x) \oplus g(x \oplus \alpha_3) = 0$ and $\Delta(\alpha_3) = 2^{2k}$. The global propagation $\sigma(g)$ is easy to compute and

$$\sigma(g) = \sum_{\alpha \in \Sigma^{2k}} \Delta^2(\alpha) = \Delta^2(0) + \sum_{j=1}^{3} \Delta^2(\alpha_j) = 4 \cdot 2^{4k} = 2^{4k+2}.$$

The collection of balanced function can be expanded by using a nonsingular linear transformation. Denote

$$g^*(x) = g(xA)$$

where A is any nonsingular $2k \times 2k$ matrix over $GF(2)$. It can be proved that the function g^* is balanced and satisfies the propagation criterion with respect to all but three non-zero vectors. The nonlinearity of g^* satisfies $N_{g^*} \geq 2^{2k-1} - 2^k$. Note that $\sigma(g)$ is invariant under any nondegenerate linear transformation on the variables. Thus $\sigma(g^*) = 2^{4k+2}$. This value compares quite favourably with the lower bound on $\sigma(h)$ which is 2^{4k}. Unfortunately, the function g has three linear structures although it satisfies the propagation criterion with respect to other nonzero vectors.

4.5.2 Concatenating Linear Functions

Assume we have two collections of Boolean variables $y = (y_1, \ldots, y_p)$ and $x = (x_1, \ldots, x_q)$ $(p < q)$. We can build up a Boolean function over Σ^{p+q} by concatenating 2^p linear functions each one over Σ^q. The collection of non-zero linear functions used in the construction is denoted by $\Re = \{\varphi_0, \ldots, \varphi_{2^p-1}\}$ and $\varphi_i \neq \varphi_j$ for any $i \neq j$, $\varphi_i : \Sigma^q \to GF(2)$. More precisely, we construct balanced, nonlinear functions by combining the linear functions from $\Re$ as follows:

$$g(z) = g(y, x) = \bigoplus_{\delta=0,\ldots,2^p-1} D_\delta(y)\varphi_\delta(x). \tag{4.5}$$

The properties of the resulting function are summarised below (the proof of the properties can be found in [275]).

PR1. The function g is balanced.

PR2. The nonlinearity of g satisfies $N_g \geq 2^{p+q-1} - 2^{q-1}$.

PR3. The function g satisfies the propagation criterion with respect to any $\gamma = (\beta, \alpha)$ with $\beta \neq 0$ where $\beta \in \Sigma^p$ and $\alpha \in \Sigma^q$.

PR4. The degree of the function g (in the algebraic normal form) can be $p + 1$ if $\Re$ is appropriately chosen.

Let ξ_δ is the sequence of φ_δ and η is the sequence of g. Clearly, from the construction, η is the concatenation of 2^p distinct ξ_δ. Note that $H_{p+q} = H_p \otimes H_q$. So each row of H_{p+q}, say L, can be represented as the Kronecker product $L = \ell' \otimes \ell''$, where ℓ' is a row of H_p and ℓ'' is a row of H_q. If $\ell' = (a_0, \ldots, a_{2^p-1})$ then $\ell' \otimes \ell'' = (a_0\ell'', \ldots, a_{2^p-1}\ell'')$ and the string $a_i\ell''$ is equal to ℓ'' if $a_i = 1$ or $-\ell''$ if $a_i = -1$. Since different rows of H_p are orthogonal, we have

$$\langle \eta, L \rangle = \begin{cases} 2^q & \text{if } f \in \Re, \text{ where } L = \ell' \otimes \ell'' \\ 0 & \text{if } f \notin \Re, \text{ where } L = \ell' \otimes \ell'' \end{cases} \tag{4.6}$$

where f is the linear function corresponding to ℓ''. There are $2^p \cdot 2^p$ different vectors $L = \ell' \otimes \ell''$ which can be constructed from 2^p linear functions from $\Re$. From Equations (4.4) and (4.5), we can obtain that the global propagation of g is

$$\sigma(g) = 2^{-p-q}2^p \cdot 2^p \cdot 2^{4q} = 2^{p+3q}.$$

The parameter $\sigma(g)$ is invariant under any nondegenerate linear transformation on the variables. Thus $\sigma(g^*) = 2^{p+3q}$ where $g^*(z) = g(Az)$. The lower bound of $\sigma(f)$, where f is a function on V_{p+q}, is 2^{2p+2q}. As we know this bound is reached only by bent functions. The nonlinearity and the number of vectors for which the propagation criterion is satisfied, are the same for both g and g^*.

The above construction applies 2^p different nonzero linear functions. There are no other restrictions imposed on the set $\Re$. We can improve the construction when we select the set $\Re$ more carefully. The

rank of the set of linear functions is the number of all linearly independent elements (functions) in the set. Assume that there is δ_0 such that the rank of the set

$$\{\varphi_\delta \oplus \varphi_{\delta_0} | \delta = 0, \ldots, 2^p - 1\} \tag{4.7}$$

is equal to q. Next we are going to show that the function g defined by Equation (4.5) has no linear structure. Consider

$$g(z) \oplus g(z \oplus \gamma) = g(y, x) \oplus g(y \oplus \beta, x \oplus \alpha) \tag{4.8}$$

As we know the function (4.8) is balanced for $\beta \neq 0$ (see the property PR3). So we can find linear structures only when $\beta = 0$. The expression (4.8) reduces to

$$\begin{aligned} g(z) \oplus g(z \oplus \gamma) &= g(y, x) \oplus g(y, x \oplus \alpha) \\ &= \bigoplus_{\delta=0,\ldots,2^p-1} D_\delta(y)(\varphi_\delta(x) \oplus \varphi(x \oplus \alpha)) = \bigoplus_{\delta=0,\ldots,2p-1} D_\delta(y)\varphi_\delta(\alpha). \end{aligned} \tag{4.9}$$

Clearly, $\gamma = (0, \alpha)$ is a linear structure if and only if (4.9) is constant or equivalently $\varphi_\delta(\alpha) = c$. This is true when

$$\varphi_\delta(\alpha) \oplus \varphi_{\delta_0}(\alpha) = 0 \tag{4.10}$$

for every $\delta = 0, \ldots, 2^p - 1$ where $c \in \Sigma$. Since the rank of $\{\varphi_\delta \oplus \varphi_{\delta_0} | \delta = 0, \ldots, 2^p - 1\} = q$, there exists no nonzero α satisfying (4.10) which is equivalent to the set of linear equations. This proves that g has no linear structures.

The condition imposed on the set $\Re$ is easy to satisfy. For example, the following collection of linear functions $h_1(x) = x_1$, $h_2(x) = x_2$, ..., $h_q(x) = x_q$ are linearly independent over Σ^q. Let φ_0 be an arbitrary linear function on Σ^q. Denote $\varphi_j = h_j \oplus \varphi_0$, $j = 1, 2, \ldots, q$. Thus $\varphi_1 \oplus \varphi_0$, ..., $\varphi_q \oplus \varphi_0$ are linearly independent. The set $\Re$ has to have 2^p linear functions. It contains the following linear functions: $\varphi_1 = h_1 \oplus \varphi_0$, ..., $\varphi_q = h_q \oplus \varphi_0$, φ_0 as the $(q+1)$-th linear function. The rest can be selected arbitrarily from the other nonzero linear functions.

4.6 S-Box Design

Single Boolean functions are basic elements which can be used to construct more complex (and useful from a cryptographic point of view) structures called *S-boxes*. An $n \times k$ S-box is a mapping from Σ^n to Σ^k and

$$S(x) = (f_1(x), \ldots, f_k(x))$$

where $n \geq k$ and $f_j : \Sigma^n \to GF(2)$.

The collection of cryptographically essential properties includes the following ones:

S1. Any nonzero linear combination of $f_1, \ldots, f_k$, i.e. $f = c_1 f_1 \oplus \cdots \oplus c_k f_k$, $(c_1, \ldots, c_k) \neq (0, \ldots, 0)$, should be balanced.

S2. Any nonzero linear combination of $f_1, \ldots, f_k$ should be highly nonlinear.

S3. Any nonzero linear combination of $f_1, \ldots, f_k$ should satisfy the SAC.

S4. The S-box $S(x)$ should be *regular*, i.e. each vector in Σ^k should happen 2^{n-k} times while x runs through Σ^n once.

S5. $S(x)$ should have a good XOR profile, i.e. $S(x) \oplus S(x \oplus \alpha)$ runs through some 2^{k-1} vectors in Σ^k each 2^{n-k+1} times while x runs through Σ^n once, but does not take on other 2^{k-1} vectors.

Observe that properties S2 and S4 are equivalent. Other properties may not hold simultaneously but a "reasonable" tradeoff can always be negotiated.

To illustrate the properties, consider a simple example. Let our S-box be the mapping from Σ^3 to Σ^3 such that

$$S(x) = (f_1(x), f_2(x), f_3(x))$$

where $f_1 = x_1 \oplus x_3 \oplus x_2x_3$, $f_2 = x_1 \oplus x_2 \oplus x_1x_2 \oplus x_2x_3$ and $f_3 = x_1x_2 \oplus x_2x_3 \oplus x_1x_3$. The properties S1-S5 can be verified as follows.

S1. Any nonzero linear combination of f_1, f_2, f_3, say $f = c_1f_1 \oplus c_2f_2 \oplus c_3f_3$, $(c_1, c_2, c_3) \neq (0, 0, 0)$, is balanced.

S2. Any nonzero linear combination f of f_1, f_2, f_3 has nonlinearity 2 i.e. $N_f \geq 2$ (the maximum for balanced functions on Σ^3).

S3. Any nonzero linear combination of f_1, f_2, f_3 satisfies the propagation criterion except for a single vector.

S4. $S(x)$ is regular as it is a permutation.

S5. $S(x)$ has a good XOR profile, i.e. $S(x) \oplus S(x \oplus \alpha)$ runs through some 2^2 vectors in Σ^3 each twice while x runs through Σ^3 once and does not take on other 2^2 vectors. More precisely, let $\alpha = (001)$. Then $S(x) \oplus S(x \oplus \alpha)$ runs through vectors (010), (011), (100), (101) twice while x runs through Σ^3 once. If $\alpha = (111)$, then $S(x) \oplus S(x \oplus \alpha)$ runs through vectors (001), (011), (101), (111) twice while x runs through Σ^3 once.

Permutations defined in $GF(2^n)$ can be searched for ones with good cryptographic properties. Pieprzyk [235] proved that exponentiation can produce cryptographically strong S-boxes. Being more specific, the S-boxes $S : \Sigma^n \to \Sigma^n$ defined as $S(x) = x^3$, $x \in GF(2^n)$ where n is odd, are permutations and they have the following properties ([235, 216, 217, 14]):

S1' Any nonzero linear combination of the co-ordinate functions, is balanced. This results from the fact that cubing is a permutation. Any nonzero linear combination f of the co-ordinate functions has a high nonlinearity and $N_f \geq 2^{n-1} - 2^{\frac{1}{2}(n-1)}$.

S3' Any nonzero linear combination of the co-ordinate functions satisfies the propagation criterion except for a single nonzero vector.

S5' $S(x)$ has a good XOR profile, i.e. $S(x) \oplus S(x \oplus \alpha)$ runs through a subset of 2^{n-1} vectors in Σ^n twice while x runs through Σ^n once. The remaining 2^{n-1} vectors do not occur.

The design of S-boxes is not free from some pitfalls. They are especially dangerous when having a cryptographically strong S-box, one would like to modify it by adding or reducing output bits. Consider the S-box $S(x) = (f_1(x), \ldots, f_k(x))$ which is regular and has a good XOR profile where $f_i : \Sigma^n \to GF(2)$ for $i = 1, \ldots, k$. It turns out (see [278]) that $S(x) = (f_1(x), \ldots, f_t(x))$ where $t < k$ is regular but does not have a good XOR profile.

On the other hand, for any regular S-box $S(x) = (f_1(x), \ldots, f_k(x))$ with a good XOR profile, there is a collection of functions $f_{k+1}(x), \ldots, f_s(x)$ such that the extended S-box $S'(x) = (f_1(x), \ldots, f_k(x), f_{k+1}(x), \ldots, f_s(x))$ is a regular mapping from Σ^n to Σ^s but does not have a good XOR profile.