Generalization of Siegenthaler Inequality and Schnorr-Vaudenay Multipermutations

Paul Camion * and Anne Canteaut **

INRIA Projet Codes
Domaine de Voluceau
78153 Le Chesnay Cedex, FRANCE
email:{Paul.Camion, Anne.Canteaut}@inria.fr

Abstract. Siegenthaler inequality shows the existence of a tradeoff between the correlation-immunity order and the nonlinearity order of a Boolean functions. We generalize this result to correlation-immune functions over any finite field. We then construct a family of correlation-immune functions achieving this bound; these functions are notably well-suited for combining linear feedback shift registers. We also apply this result to the cryptanalysis of any cryptographic primitive based on boxes connected by a network. Schnorr and Vaudenay have previously recommended that these boxes should be multipermutations; we here refine this condition since we show that each binary component of these multipermutations, seen as a Boolean function, should have low degree.

KEYWORDS: correlation-immune functions, stream ciphers, hash functions, cryptanalysis, orthogonal arrays.

1 Introduction

Correlation-immune functions were first introduced by Siegenthaler as a class of suitable functions for combining the outputs of several linear feedback shift registers: they lead to the construction of running-key generators for stream ciphers which resist the correlation attack [12]. Several other applications afterwards emerged and correlation-immune functions (or resilient functions) are now preponderant objects in cryptography. Schnorr and Vaudenay [10] notably pointed out their importance in the design of conventional cryptographic primitives, such as hash functions. Maurer and Massey [6] developed the related concept of perfect local randomizer for constructing additive stream ciphers which are provably-secure under the restriction that the enemy can only obtain a limited number of plaintext digits. Both of these applications often consider correlation-immune functions over finite fields since most of the cryptographic primitives do not deal

* Centre National de la Recherche Scientifique
** Grant-holder from the DRET, also with École Nationale Supérieure de Techniques Avancées, 32 boulevard Victor, F-75015 Paris.

N. Koblitz (Ed.): Advances in Cryptology - CRYPTO '96, LNCS 1109, pp. 372-386, 1996.

with bits but with m-bit words according to the characteristics of the used electronic components. Nevertheless considering the correlation-immunity order of a function is usually not sufficient for obtaining good cryptographic properties. For designing good running-key generators, for example, the Berlekamp-Massey shift register synthesis also requires the combining function to be non linear. Unfortunately this condition can be incompatible with a high correlation-immunity order since Siegenthaler proved in [12] the existence of a tradeoff between the correlation-immunity order and the nonlinearity order of Boolean functions. We here generalize Siegenthaler's work as we give a similar result for correlation-immune functions over any field.

Section 2 briefly recalls the equivalence between correlation-immune functions over any finite alphabet and orthogonal arrays. Then we study the properties of the algebraic normal form of correlation-immune functions over a finite field and we give an upper bound on their nonlinearity, notion that has to be defined. We then construct in section 4 a family of t-resilient functions with optimal nonlinearity over some finite fields, which can be used for combining LFSRs. We exhibit in section 5 a similar result for q-ary functions which are correlation-immune with respect to $\mathbf{F}_{q^k}$. Section 6 applies these results to the functions used for designing cryptographic primitives: Schnorr and Vaudenay give in [10] a general method for cryptanalyzing any cryptographic primitive based on boxes connected by a network; they therefore recommend that all the boxes should be multipermutations over $\mathbf{F}_{2^m}$. We here show that every binary components of these multipermutations, considered as Boolean functions, should have low degree.

2 Correlation immune functions and orthogonal arrays

Let $\mathcal{F}$ denote a finite alphabet with q elements ($q \geq 2$) and let E be a finite set. Let $f : \mathcal{F}^n \to E$ be a function and let $\{X_1, X_2, ..., X_n\}$ be a set of random input variables assuming values from $\mathcal{F}$ with independent equiprobable distributions (*i.e.* every input vector occurs with probability $\frac{1}{q^n}$).

The function f may satisfy the following properties:

- f is *balanced* if its output is uniformly distributed.
- f is *correlation-immune with respect to the subset* $T \subset \{1, 2, \ldots, n\}$ if the probability distribution of the output is unaltered when $\{X_i, i \in T\}$ is fixed and $\{X_i, i \notin T\}$ is a set of independent equiprobable random variables.
- f is *t-th order correlation-immune* if for every T of cardinality at most t, f is correlation-immune with respect to T.
- f is *t-resilient* if f is t-th order correlation-immune and balanced.

Correlation-immune functions are closely related to the combinatorial structures introduced by Rao as orthogonal arrays [7].

Definition 1. An orthogonal array A of size m, n constraints, strength t and index λ over the alphabet $\mathcal{F}$ (or with q levels) is an $m \times n$ array of which rows are the vectors from a subset M of $\mathcal{F}^n$ such that $|M| = m$ which has the property

that in any subset of t columns of A, each of the q^t vectors of $\mathcal{F}^t$ appears exactly λ times as a row. Such an array is denoted by (m, n, q, t). Clearly $m = \lambda q^t$.

The equivalence between correlation-immune functions and orthogonal arrays was first proved in [3] for the Boolean case, and in [4] for functions over any finite field. In fact characterizing the t-th order correlation immune functions in terms of orthogonal arrays is merely translating the probability definition into an enumeration definition. This characterization then holds for any finite alphabet $\mathcal{F}$.

Proposition 2. *If the independent equiprobable random variables $X_1, \ldots, X_n$ are defined on a finite alphabet $\mathcal{F}$, then $f(X_1, \ldots, X_n)$, which has its values in a finite set E, is a t-th order correlation immune function with respect to the alphabet $\mathcal{F}$ if and only if $\forall y \in E$, $f^{-1}(y)$ consists in the rows of an orthogonal array of strength t over $\mathcal{F}$.*

Additionally, f is t-resilient if $\forall y, y' \in E, |f^{-1}(y)| = |f^{-1}(y')|$.

3 Algebraic normal form and nonlinear order of correlation immune functions over any finite field

Parallel to the order of correlation-immunity, the order of nonlinearity of a function is a fundamental parameter in cryptography. In particular when the function is used for combining the outputs of some linear feedback shift registers, it completely determines the linear complexity of the resulting running-key generator (see [8, 5]). For Boolean functions, this nonlinearity order is directly obtained from the algebraic normal form of the function as the degree of the corresponding binary polynomial. Such a canonical expression can also be defined for any function f from $\mathcal{F}^n$ to $\mathcal{F}^\ell$, if $\mathcal{F}$ is the finite field $GF(q)$. In this general situation we first introduce through Theorem 7 the notion of *optimal nonlinearity*. From now on this field is be denoted by $\mathbf{F}_q$ and $\mathcal{F}^\ell$ is identified with the finite field $\mathbf{F}_{q^\ell}$.

Notation 3 *Let $\mathcal{M}_n$ be the algebra $\mathbf{F}_{q^\ell}[x_1, \ldots, x_n]/(x_1^q - x_1, \ldots, x_n^q - x_n)$. For a polynomial $\theta \in \mathcal{M}_n$ which is considered the smallest degree multivariate polynomial in its class, we denote by $Deg_{x_i}\ \theta$ the degree of θ in the variable x_i, $i = 1, \ldots, n$.*

Definition 4. In the algebra $\mathcal{M}_n$ we call $L_\alpha(x_i) = \prod_{\substack{\beta \in \mathbf{F}_q \\ \beta \neq \alpha}} (x_i - \beta)$, $\alpha \in \mathbf{F}_q$ the Lagrange univariate idempotents (with respect to x_i).

The following known lemma which is easily proven by induction on n enables us to show the existence and unicity of the algebraic normal form for any function $f : \mathbf{F}_q^n \to \mathbf{F}_{q^\ell}$.

Lemma 5. *Let $\theta \in \mathcal{M}_n$ be considered as a polynomial function defined on $\mathbf{F}_q^n$. If θ vanishes for all $x \in \mathbf{F}_q^n$, then all its coefficients are zero.*

Theorem 6. *For any function $f : \mathbf{F}_q^n \to \mathbf{F}_{q^\ell}$ there exists a unique polynomial function $\theta \in \mathcal{M}_n$ (thus with $Deg_{x_i}\ \theta \leq q-1$, $i = 1, \ldots, n$) such that, for all x in $\mathbf{F}_q^n$, $f(x) = \theta(x)$. This polynomial θ is called the* algebraic normal form *of f.*

The proof of this theorem relies on Lagrange interpolation since we can write $\theta(x) = \sum_{\alpha \in \mathbf{F}_q} -f(\alpha) L_\alpha(x)$.

Using a combining function which has both highest possible nonlinearity and correlation-immunity orders would be suitable in order to protect the resulting running-key generator from all known attacks. Unfortunately, there exits a tradeoff between these parameters: a low nonlinearity is the price to pay for a high correlation-immunity order. Siegenthaler proved in [12] that for any Boolean function f from $\mathbf{F}_2^n$ to $\mathbf{F}_2$, the nonlinearity order d and the correlation-immunity order t always satisfy $d + t \leq n$. We here exhibit a similar relation for any function from $\mathbf{F}_q^n$ to $\mathbf{F}_{q^\ell}$. Actually those relations for $q > 2$ are derived from stronger properties.

Theorem 7. *Let there be given a function $f : \mathbf{F}_q^n \to \mathbf{F}_{q^\ell}$. If f is t-th order correlation-immune (resp. is t-resilient) with respect to $\mathbf{F}_q$, then for every monomial μ in the algebraic normal form θ of f there exists a subset T of $\{1, \ldots, n\}$ of size t (resp. of size $t+1$ provided $q^\ell \neq 2$ or $n \neq \ell + t$) such that $Deg_{x_i}\ \mu \leq q-2$, $\forall i \in T$.*

Proof. Let j variables be fixed amongst $x_1, \ldots, x_n$, for example and without loss of generality we choose $x_{n-j+1} = x^*_{n-j+1}, \ldots, x_n = x^*_n$. Then we have

$$\theta(x_1, \ldots, x_{n-j}, x^*_{n-j+1}, \ldots, x^*_n) = \sum_{\alpha \in \mathbf{F}_q^{n-j}} -f(\alpha_1, \ldots, \alpha_{n-j}, x^*_{n-j+1}, \ldots, x^*_n) \prod_{i=1}^{n-j} L_{\alpha_i}(x_i) \qquad (1)$$

Let $I_{1,\ldots,n-j}(v) = \{\alpha \in \mathbf{F}_q^{n-j},\ f(\alpha_1, \ldots, \alpha_{n-j}, x^*_{n-j+1}, \ldots, x^*_n) = v\}$.
If f is t-th order correlation-immune, then

$$\forall j \leq t, \forall v \in \mathbf{F}_{q^\ell},\ \ |I_{1,\ldots,n-j}(v)| = \frac{|f^{-1}(v)|}{q^j}$$

Thus $\forall j \leq t, \forall v \in \mathbf{F}_{q^\ell},\ \ |I_{1,\ldots,n-j}(v)| = q^{t-j} |I_{1,\ldots,n-t}(v)|$. This relation shows that for all $j < t$, $|I_{1,\ldots,n-j}(v)| = 0$ in $\mathbf{F}_{q^\ell}$. Then, for all $j < t$, relation 1 can be written as:

$$\theta(x_1, \ldots, x_{n-j}, x^*_{n-j+1}, \ldots, x^*_n) = \sum_{v \in \mathbf{F}_q^\ell} -v \sum_{\alpha \in I_{1,\ldots,n-j}(v)} \prod_{i=1}^{n-j} L_{\alpha_i}(x_i)$$

Since each L_{α_i} is monic polynomial of degree $q-1$, the coefficient of degree $(q-1)(n-j)$ of $\sum_{\alpha \in I_{1,\ldots,n-j}(v)} \prod_{i=1}^{n-j} L_{\alpha_i}(x_i)$ is equal to $|I_{1,\ldots,n-j}(v)|$ which is zero in $\mathbf{F}_{q^\ell}$.

Since this is true for any other choice of j variables, $j < t$, amongst $x_1, \ldots, x_n$ it ensures that any monomial of θ contains no product of $n-t+1$ or more variables having simultaneously degree $q-1$, as asserted.

Furthermore if f is balanced, we have $\forall v \in \mathbf{F}_{q^\ell}, |f^{-1}(v)| = q^{n-\ell}$. Thus we additionally have:

$$\forall v \in \mathbf{F}_{q^\ell},\ |I_{1,\ldots,n-t}(v)| = q^{n-\ell-t}$$

$|I_{1,\ldots,n-t}(v)|$ is then zero in $\mathbf{F}_{q^\ell}$ provided $n-t-\ell > 0$.
Moreover if $n = t+\ell$ then $|I_{1,\ldots,n-t}(v)| = 1$. The coefficient of a monomial μ of θ such that $Deg_{x_i}\ \mu = q-1,\ i = 1, \ldots, n-j$ is equal to $\sum_{v \in \mathbf{F}_{q^\ell}} -v|I_{1,\ldots,n-t}(v)| = \sum_{v \in \mathbf{F}_{q^\ell}} -v = 0$ provided $q^\ell \neq 2$, as asserted. This means in particular that, if $n \neq t+\ell$ or $q^\ell \neq 2$, θ contains no product of $n-t$ or more variables having simultaneously degree $q-1$.

Remark. The previous proof also implies a stronger condition on the algebraic normal form of some t-th order correlation immune functions, even if they are not balanced: if $f : \mathbf{F}_q^n \to \mathbf{F}_{q^\ell}$ is a t-th order correlation immune function with respect to $\mathbf{F}_q$ such that:

$$\forall v \in \mathbf{F}_{q^\ell},\ \frac{|f^{-1}(v)|}{q^t} = 0 \bmod q$$

then the assertion of the theorem on balanced functions holds.

We can then deduce from this theorem an inequality which generalizes Siegenthaler's one.

Corollary 8. *Let $f : \mathbf{F}_q^n \to \mathbf{F}_{q^\ell}$ be a t-th order correlation-immune function with respect to $\mathbf{F}_q$. Then the total degree d of its algebraic normal form satisfies*

$$d + t \leq (q-1)n$$

If f is additionally balanced and $n \neq \ell + t$ or $q^\ell \neq 2$, then

$$d + t \leq (q-1)n - 1$$

Definition 9. A function $f : \mathbf{F}_q^n \to \mathbf{F}_{q^\ell}$ either correlation immune of order t or t-resilient has *optimal nonlinearity* if, for its algebraic normal form θ, all corresponding bounds in Theorem 7 are tight.

Optimal nonlinearity of resilient functions used for combining LFSRs will permit to achieve high linear complexity as shown in the following section.

4 Construction of t-resilient functions with optimal nonlinearity over any finite field

We here construct t-resilient functions $f : \mathbf{F}_q^n \to \mathbf{F}_q$ with optimal nonlinearity. Such functions are especially useful for designing running-key generators by combining linear feedback shift registers. The combining function has indeed to be correlation-immune and balanced since the output digits have to be uniformly distributed.

4.1 Combining LFSRS

We first recall some well-known results which show that optimal nonlinearity leads to a maximal linear complexity of the resulting running-key generator.

Let **a** denote the sequence given by

$$a_{m+i} = -b_{m-1}a_{m+i-1} - b_{m-2}a_{m+i-2} - \ldots - b_0 a_i \qquad (2)$$

A polynomial $g(X) = b_0 + b_1 X + \ldots + b_{m-1}X^{m-1} + X^m \in \mathbf{F}_q[X]$ such that a sequence **a** verifies (2) is called a *recurrence polynomial* (or *characteristic polynomial*) for **a**. Whenever a polynomial such as g is a recurrence polynomial for **a**, we say that **a** is driven by g.

For a sequence **a** we denote by $L(\mathbf{a})$ its *linear complexity* which is the smallest degree of a recurrence polynomial which drives **a**. There clearly is a unique monic polynomial with degree $L(\mathbf{a})$ which drives a sequence **a**. It is called the *minimal polynomial* of the sequence. The period of the sequence is then equal to the order of its minimal polynomial.

But, even if the recurrence polynomial is properly chosen, the linear complexity of the sequence is often smaller as we wish. A well-known method for increasing it consists in using several LFSRs with different feedback polynomials. Their output sequences are then taken as arguments of the algebraic normal form of a combining function $f : \mathbf{F}_q^k \to \mathbf{F}_q$ whose output then forms the running-key, as depicted in Figure 1

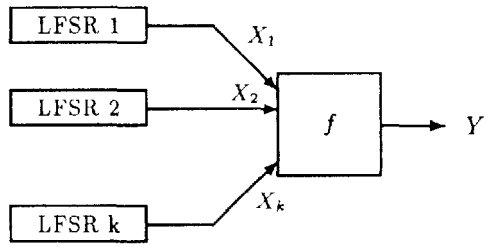

Fig. 1. Combining LFSRs

An estimation of the linear complexity of this resulting generator can be obtained thanks to the following theorems proved in [9, 5, 1].

Proposition 10. *Let* **a** *and* **b** *be two sequences whose minimal polynomials are respectively* g_1 *and* g_2. *Then the linear complexity of the sum sequence satisfies* $L(\mathbf{a}+\mathbf{b}) \leq L(\mathbf{a}) + L(\mathbf{b})$ *where equality holds if and only if* $gcd(g_1, g_2) = 1$.

We now compute the linear complexity of a product of sequences whose respective minimal polynomials are homogeneous polynomials *i.e.* products of distinct irreducible polynomials with equal degree.

Proposition 11. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be two sequences whose respective minimal polynomials* g_1 *and* g_2 *are homogeneous polynomials with respective orders* t_1 *and* t_2. *The Hadamard product sequence* $\mathbf{ab} = (a_n b_n)_{n \geq 0}$ *is then driven by an homogeneous polynomial whose order divides* $lcm(t_1, t_2)$ *and its linear complexity is at most* $L(\mathbf{a})L(\mathbf{b})$. *Moreover, if* $gcd(t_1, t_2) = 1$, *then* $L(\mathbf{ab}) = L(\mathbf{a})L(\mathbf{b})$.

We now come to Herlestam's result [5] on the linear complexity of a power sequence.

Theorem 12. *If* $0 \leq s < q$, $q = p^e$, $s = \sum_{i=0}^{e-1} s_i p^i$, $0 \leq s_i < p$, *then*

$$L(\mathbf{a}^s) \leq \prod_i \binom{L(\mathbf{a}) + s_i - 1}{s_i}$$

with equality if $\mathbf{a}$ *is a ML-sequence. In particular, when* $p = 2$,

$$L(\mathbf{a}^s) \leq L(\mathbf{a})^{w(s)}$$

where $w(s)$ *is the Hamming weight of* s *and where equality holds if* $\mathbf{a}$ *is a ML-sequence.*

Equality in the ML-case is due to [1]. The proof relies on a corollary of Lucas Theorem for multinomial coefficients.

For a product sequence the best that can be expected for the linear complexity is that of the sequences being multiplied. Whenever a product sequence satisfies this upper bound with equality, one says that this product sequence attains *maximum linear complexity*. We thus see that the optimal nonlinearity of a t-resilient function leads to maximum linear complexity for the resulting pseudo-random generator.

4.2 Construction of resilient functions with optimal non-linearity

Lemma 13. *Let* $\mathcal{M}$ *be the algebra* $\mathbf{F}_q[z]/(z^q - z)$.
We have in $\mathcal{M}$ *that* $0 \leq degree(z^{i(q-2)}) < q - 2$ *for all* $2 \leq i < q - 1$, *and for even* q, $1 \leq degree(z^{j\frac{q-2}{2}}) < q - 2$ *for all* $3 \leq j \leq q - 2$.

Proof. Indeed we first have $z^{i(q-2)} = z^{q+q(i-1)-2i} = z^q z^{i-1-2i} = z^{q-i-1}$, with $0 \leq q - i - 1 < q - 2$.
Next we write $j = 2a + b$ with $b \in \{0, 1\}$. If $b = 0$, we make $i = a$ above since $2 \leq a \leq q - 2$. If $b = 1$, we have $1 \leq a < \frac{q-2}{2}$. Then we obtain $Deg_z^{j\frac{q-2}{2}} = Deg_z\, z^{a(q-2)} z^{\frac{q-2}{2}} = \frac{q-2-2a}{2} < q - 2$.

We here essentially construct (q^t, n, q, t) orthogonal arrays of index unity.

Proposition 14. *For all* $q = p^a$ *with* $p \neq 3$ *and* $q \neq 4$ *there exists an* 1*-resilient function* $f : \mathbf{F}_q^2 \to \mathbf{F}_q$ *with optimal nonlinearity. With the additional assumption on* q *that* $q \not\equiv 1 \bmod 3$ *we have that there exist a* $(n-1)$*-resilient function* $f : \mathbf{F}_q^n \to \mathbf{F}_q$ *with optimal nonlinearity for every* n *if* q *is even and for every odd* n *if* q *is odd.*

Proof. Proof of the first assertion.
For odd characteristic $p > 3$, we define $f(x, y) = (x^{q-2} + y^{q-2} + 1)^{q-2}$, and for even $q > 4$, $f(x, y) = (x^{q-2} + y^{\frac{q-2}{2}} + 1)^{q-5}$. In both cases f is 1-resilient since $gcd(q-2, q-1) = 1$ and for even $q > 4$, $gcd(\frac{q}{2}-1, q-1) = 1$, $gcd(q-5, q-1) = 1$, which shows that all these exponentiations permute the finite field $\mathbf{F}_q$. In view of Lemma 13 we point out that, in the first case, the coefficient of $x^{q-2}y^{q-2}$ is $(q-2)(q-3)$ which is not a multiple of $p > 3$. In the second case we see that the coefficient of $x^{q-2}y^{q-2}$ is $3\binom{q-5}{3} \equiv 1 \bmod 2$.
Proof of the second assertion, by induction on n.

- q odd. We have that $g_3(x_1, x_2, x_3) = (x_1^{q-2} + x_2^{q-2} + x_3^{q-2})^3$ which contains the monomial $6(x_1x_2x_3)^{q-2}$.

Next we define $g_{2r+1}(x_1, \ldots, x_{2r+1}) = (g_{2r-1}(x_1, \ldots, x_{2r-1}) + x_{2r}^{q-2} + x_{2r+1}^{q-2})^3$ containing the term $6g_{2r-1}(x_1, \ldots, x_{2r-1})x_{2r}^{q-2}x_{2r+1}^{q-2}$.

- q even. We first have $g_2(x_1, x_2) = (x_1^{\frac{q}{2}-1} + x_2^{q-2})^3$ where only $3x_1^{q-2}x_2^{q-2}$ has x_2^{q-2} as a factor. Then $g_{n+1}(x_1, \ldots, x_{n+1}) = (g_n(x_1, \ldots, x_n) + x_{n+1}^{\frac{q}{2}-1})^3$.

In that last polynomial, the only terms with x_{n+1}^{q-2} as a factor are in $3g_n(x_1, \ldots, x_n)x_{n+1}^{q-2}$. All these functions g_n are $n-1$-resilient since $gcd(3, q-1) = 1$ by assumption, and $gcd(q-2, q-1) = 1$. By Theorem 7, they have optimal nonlinearity. This still holds for $q = 2$: $f(x_1, \ldots, x_n) = x_1 + \ldots + x_n$ is an $(n-1)$-resilient function with optimal nonlinearity since $1 + t = n$.

Proposition 15. *Let $q \neq 2$ or $t \neq n-1$. Let $f_1, f_2 : \mathbf{F}_q^n \to \mathbf{F}_q$ be two t-resilient functions with optimal nonlinearity , such that $degree(f_1 - f_2) = degree(f_1)$. Then $g : \mathbf{F}_q^{n+1} \to \mathbf{F}_q$ defined by*

$$g(x_1, \ldots, x_{n+1}) = x_{n+1}^{q-1} f_1(x_1, \ldots, x_n) + (1 - x_{n+1}^{q-1}) f_2(x_1, \ldots, x_n)$$

is a t-resilient function with optimal nonlinearity.

From both previous propositions we deduce the following theorem:

Theorem 16. *Let $q = p^a$ with $p \neq 3$ and $q \neq 4$. For all $n > 1$, there exits a 1-resilient function $f : \mathbf{F}_q^n \to \mathbf{F}_q$ with optimal nonlinearity. If $q \not\equiv 1 \bmod 3$, then for all $n > 1$, there exits a t-resilient function $f : \mathbf{F}_q^n \to \mathbf{F}_q$ with optimal nonlinearity for all $t < n$ if q is even, and for all even $t < n$ if q is odd.*

Proof. By Proposition 14, if t agrees with the above assumptions, there exists a t-resilient function $g : \mathbf{F}_q^{t+1} \to \mathbf{F}_q$ with optimal nonlinearity. Applying Proposition 15 with $f_1 = g$ and $f_2 = \alpha g$, where $\alpha \in \mathbf{F}_q \setminus \{0, 1\}$ leads to a t-resilient function with $t + 2$ variables and optimal nonlinearity . If we iterate this construction $n - t - 1$ times, we obtain a t-resilient function with n variables and optimal nonlinearity. Siegenthaler proved this result in Boolean case.

Example 1. We here construct a 2-resilient function with 4 variables over $\mathbf{F}_8$. Proposition 14 enables us to construct functions g_1 and g_2 which are respectively 1-resilient with 2 variables and 2-resilient with 3 variables. Both normal forms have optimal nonlinearity:

$g_1(x_1,x_2) = (x_1^6 + x_2^3)^3 = x_1^6x_2^6 + x_1^5x_2^3 + x_1^4 + x_2^2$
$g_2(x_1,x_2,x_3) = (g_1(x_1,x_2)+x_3^3)^3 = x_1^6x_2^6x_3^6+x_1^5x_2^3x_3^6+x_1^4x_2^5x_3^4+x_1^5x_2^7+x_1^4x_2^7x_3+$
$x_1^2x_2^6x_3^4 + x_1^7x_2^2x_3 + x_1^4x_3^6 + x_1^6x_2^3 + x_1^5x_2^3x_3 + x_1^4x_2^4 + x_1^2x_2^4x_3^2 + x_2^2x_3^6 + x_2^4x_3^3 + x_2^6 +$
$x_1^4x_3 + x_1^2x_2x_3 + x_1x_2x_3^2 + x_3^4 + x_2^2x_3 + x_3^2$

We now apply lemma 15 with $f_1 = g_2$ and $f_2 = \alpha g_2$ where $\alpha \in \mathbf{F}_8 \setminus \{0,1\}$. We then obtain a 2-resilient function f with 4-variables and optimal nonlinearity $d = 25$. The function $f(x_1,x_2,x_3,x_4)$ has 42 terms among which $(\alpha+1)x_1^6x_2^6x_3^6x_4^7$. Since this monomial contains $3 = t+1$ variables with degree $q-2$ and one with degree $q-1$, it hes optimal nonlinearity according to Theorem 7.

5 Algebraic Normal Form of q-ary functions which are correlation immune with respect to $\mathbf{F}_q^k$

We now give a new bound for the optimal nonlinearity of any function f from $(\mathbf{F}_{q^k})^n$ to $\mathbf{F}_q$ which is correlation immune with respect to $\mathbf{F}_{q^k}$.

For a polynomial $f \in \mathbf{F}_q[x_{i,j}, 1 \le i \le n,\ 0 \le j \le k-1]$, $Pdeg_{x_i}\ f$ denotes the partial degree of f in the variables $x_{i,1},\ldots,x_{i,k-1}$

Theorem 17. *Let there be given a function $f : (\mathbf{F}_{q^k})^n \to \mathbf{F}_q$ where $k > 1$. Its normal form is then a polynomial $\theta \in \mathbf{F}_q[x_{i,j}, 1 \le i \le n,\ 0 \le j \le k-1]/(x_{i,j}^q - x_{i,j})$.*
If f is t-th order correlation immune (resp. t-resilient) with respect to $\mathbf{F}_{q^k}$, then to every monomial ρ of θ corresponds a subset T of $\{1,\ldots,n\}$ of size t (resp. $t+1$) such that $Pdeg_{x_i}\rho \le (k-1)(q-1)+q-2$, $\forall i \in T$ and $Pdeg_{x_i}\rho \le k(q-1)$, $\forall i \notin T$.

Proof. Let us consider f as a function from $(\mathbf{F}_{q^k})^n$ to $\mathbf{F}_{q^k}$. Then its normal form is a polynomial $\mu \in \mathbf{F}_{q^k}[x_1,\ldots,x_n]/(x_i^{q^k} - x_i)$. Let α be a primitive element in $\mathbf{F}_{q^k}$. Then $\mathbf{F}_{q^k} = \mathbf{F}_q + \alpha\mathbf{F}_q + \ldots + \alpha^{k-1}\mathbf{F}_q$ and to any $x_i \in \mathbf{F}_{q^k}$ can be associated a polynomial of $\mathbf{F}_{q^k}[x_{i,j},\ 0 \le j \le k-1]$, linear in each indeterminate.
The function f can therefore be written as a polynomial θ in the algebra $\mathbf{F}_{q^k}[x_{i,j}]$ modulo the ideal generated by $x_{i,j}^q - x_{i,j}$, $1 \le i \le n$, $0 \le j \le k-1$. If f takes its values in $\mathbf{F}_q$, then $\theta^q(x) = \theta(x)$ for all $x \in \mathbf{F}_q^{kn}$. Thus $\theta = \theta^q$ and all coefficients of θ lie in $\mathbf{F}_q$ in view of Lemma 5.

We now write x_i^s for all $s < q^k$ as a polynomial in $x_{i,0},\ldots,x_{i,k-1}$. Let $s = s_0 + s_1q + \ldots + s_{k-1}q^{k-1}$ be the q-ary decomposition of s and $w_q(s) = \sum_{i=0}^{k-1} s_i$. Then we have:

$$x_i^s = \prod_{j=0}^{k-1}(x_{i,0} + \alpha x_{i,1} + \ldots + \alpha^{k-1}x_{i,k-1})^{s_jq^j}$$
$$= \prod_{j=0}^{k-1}(x_{i,0} + \alpha^{q^j}x_{i,1} + \ldots + \alpha^{(k-1)q^j}x_{i,k-1})^{s_j}$$

Then $Pdeg_{x_i}(x_i^{s_i})$ is at most $w_q(s)$. Let $\prod_{i=1}^{n} x_i^{s_i}$ be a monomial of θ. By Theorem 7, if f is t-th order correlation immune with respect to $\mathbf{F}_{q^k}$, this product contains at most $n-t$ variables of degree q^k-1. Each of these variables gives a polynomial over $\mathbf{F}_q$ of degree at most $k(q-1)$. The degree s of any other variable in the product is less than or equal to q^k-2; then $w_q(s) \leq (k-1)(q-1)+q-2$. In other words there exists a $T \subset \{1,\ldots,n\}$, $|T| = t$ such that $Pdeg_{x_i}\rho \leq (k-1)(q-1)+q-2$, $\forall i \in T$.

If f is additionally balanced, we have for all $v \in \mathbf{F}_q, |f^{-1}(v)| = q^{nk-1}$. Since $k > 1$ and $t < n$, $\frac{|f^{-1}(v)|}{q^{kt}} = 0 \bmod q$. In view of the remark following Theorem 7, we then obtain the expected result.

Remark. As for Theorem 7 the strongest condition may hold even if f is not balanced. A necessary condition for having the property asserted for balanced functions is:

$$\forall v \in \mathbf{F}_q, \quad \frac{|f^{-1}(v)|}{q^{kt}} = 0 \bmod q$$

Corollary 18. *Let $f : \mathbf{F}_{q^k}^n \to \mathbf{F}_q$ be a t-th order correlation-immune function with respect to $\mathbf{F}_{q^k}$. Then the total degree d of its algebraic normal form satisfies*

$$d + t \leq (q-1)kn$$

If f is additionally balanced and $k > 1$, then

$$d + t \leq (q-1)kn - 1$$

Example 2.
Let ϕ: $\mathbf{F}_8 \times \mathbf{F}_8 \to \mathbf{F}_8$
$(x;y) \mapsto (x^3+y^3)^3$

Let α be a root of X^3+X+1. To each element x in $\mathbf{F}_8$ we associate the polynomial $x_0+\alpha x_1+\alpha^2 x_2$ and we now consider ϕ as a function from $\mathbf{F}_2^6$ to $\mathbf{F}_2^3$. Each of its components f_0, f_1, f_2 defined by $\phi = f_0 + \alpha f_1 + \alpha^2 f_2$ is a Boolean function with 6 Boolean variables and it is obviously 1-resilient with respect to $\mathbf{F}_8$. All of them have optimal nonlinearity; the normal form of f_0 is for example: $f_0(x_0;x_1;x_2;y_0;y_1;y_2) = x_0 + y_0 + x_1y_2 + x_2y_1 + x_0x_1y_1 + x_1y_0y_1 + x_2y_0y_1 + x_0x_1y_2y_2 + x_0x_2y_2 + x_0x_2y_0y_1 + x_0x_1y_0y_2$.

6 Application to the design of cryptographic primitives

Many "conventional" cryptographic primitives consist of some small boxes connected by a graph structure. In [10] Schnorr and Vaudenay exposed a cryptanalysis method for such primitives only based on the graph structure, called the *black box cryptanalysis*. In order to maximize the complexity of this attack they recommend that all the boxes should be functions realizing perfect diffusion, and they introduce the concept of multipermutations.

Definition 19. A (r, n) multipermutation over a finite alphabet $\mathcal{F}$ is a function π from $\mathcal{F}^r$ to $\mathcal{F}^n$ such that 2 different $(r+n)$-tuples of the form $(x, \pi(x))$ cannot collide in any r positions.

All the boxes of the network representing a cryptographic primitives should then be multipermutations. Let us consider a box with r inputs and n outputs in the network. If this box is an (r, n) multipermutation then the knowledge of any $r-1$ or less words amongst inputs and outputs does not permit to determine any of the other inputs or outputs. Moreover we assume that all inputs and outputs can be deduced from the knowledge of any r of them. We will say that such a box has degree of freedom r. This means that, if you want to resolve this box (*i.e.* to find all its inputs and output) such that, for example, the first output has a given value, then you have to try all possible values for $r-1$ inputs/outputs. The complexity of the resolution is then the size of the examined space, *i.e.* $|\mathcal{F}|^{r-1}$.

In [10] Schnorr and Vaudenay applied the black box cryptanalysis for inverting or finding collisions for hash functions whose compression function is based on FFT-networks. These functions were improved in [11].
The iterative hash function $h_{k,s}$ is defined as follows: the message M is split into n blocks $M^1, \ldots, M^n$ of 2^{m+k-1} bits. For $1 \leq i \leq n$, we iterate the compression function $g_{k,s}$: $H^i = g_{k,s}(H^{i-1}, M^i)$, where H^0 is a fixed initial value. The hash value of M is then H^n. The compression function $g_{k,s}$ over $\mathbf{F}_{2^m}$ has 2^k inputs and 2^{k-1} outputs in $\mathbf{F}_{2^m}$; it is based on the FFT-network with $s+1$ layers. Then it contains $2^{k-1}(s+1)$ boxes performing a (2,2)-multipermutation over $\mathbf{F}_{2^m}$. Figure 2 gives for example the structure of $g_{2,1}$.

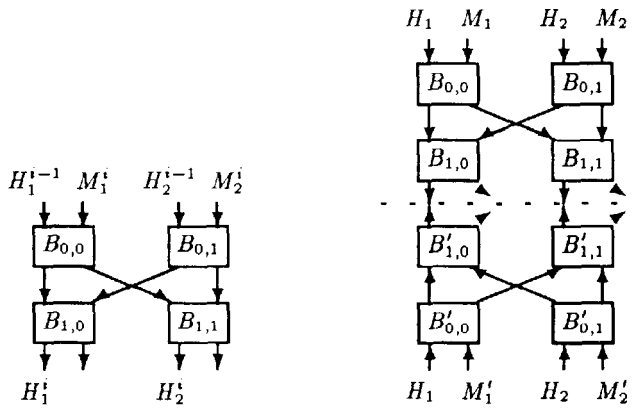

Fig. 2. The $g_{2,1}$ compression function and the corresponding collision network

It is then possible to find collisions for $h_{2,1}$ *i.e.* pairs of messages (M, M') such that $g_{2,1}(H, M) = g_{2,1}(H, M')$ by resolving the network given in Figure 2 for fixed values H_1 and H_2 in $\mathbf{F}_{2^m}$. Since all boxes are multipermutations over

$\mathbf{F}_{2^m}$, the complexity of the resolution is $(2^m)^3$ because we have to try all possible values for M_1 and M_2 for resolving boxes $B_{0,0}, B_{0,1}, B_{1,0}, B_{1,1}$. Then considering all possible values for M_1' enables us to successively resolve boxes $B_{0,0}', B_{1,0}', B_{1,1}'$ and $B_{0,1}'$.

However this collision network may be resolved with a lower complexity if inputs and outputs are not considered as elements of $\mathbf{F}_{2^m}$ anymore, but as binary strings; the degree of freedom is not preserved if the boxes are viewed as functions from $\mathbf{F}_2^{2m}$ to $\mathbf{F}_2^{2m}$. We consider for instance the $g_{2,1}$ function over $\mathbf{F}_4$ with boxes $B_{0,0}$ and $B_{1,0}$ defined by:

$$\begin{array}{rccc} \pi_1: & \mathbf{F}_2^4 & \rightarrow & \mathbf{F}_2^4 \\ & (x_1, x_0, y_1, y_0) & \mapsto & (x_0 + y_0, x_1 + y_1, x_1 + x_0 + y_1 + 1, x_1 + y_0) \end{array}$$

and boxes $B_{0,1}$ and $B_{1,1}$ defined by:

$$\begin{array}{rccc} \pi_2: & \mathbf{F}_2^4 & \rightarrow & \mathbf{F}_2^4 \\ & (x_1, x_0, y_1, y_0) & \mapsto & (x_1 + y_1, x_0 + y_0, x_1 + x_0 + y_1 + 1, x_1 + y_0) \end{array}$$

These functions are both (2,2) multipermutations over $\mathbf{F}_4$ since they correspond to pairs of orthogonal Latin squares.

We now want to obtain all the possible solutions of the collision network for fixed (H_1, H_2). We refer the reader to Figure 3.

1. Considering all possible values for M_1 and for the low-weight bit of M_2 enables us to resolve $B_{0,0}$ and to find the low-weight bit of both outputs of $B_{0,1}$. It follows that we similarly get the high-weight (resp. low-weight) bit of the left output of $B_{1,0}$ (resp. $B_{1,1}$). We now consider all possible values for the high-weight bit of M_1' and we obtain the low-weight bit of the left output and the high-weight bit of the right output of $B_{0,0}'$.
2. The set of bits for which all possible values are examined is then $E = \{$ both bits of $M_1, m_2, m_1'\}$ (these bits are underlined in Figure 3). At this step, the complexity is then $2^{|E|} = 2^4$. We now know the low-weight bit of the left input and the high-weight bit of the left output of $B_{1,0}'$. Then we deduce the low-weight bit of its right input thanks to the particular structure of the box. This leads to the knowledge of the low-weight bit of M_2' and of the right output of $B_{0,1}'$.
3. If we now examine all possible values for the high-weight bit of M_2', we completely resolve $B_{0,1}'$. The set E is now $\{$ both bits of $M_1, m_2, m_1', n_2'\}$; the complexity is 2^5. Then the structure of the box enables us to completely resolve $B_{1,1}'$ from the knowledge of its right input, the high-weight bit of its left input and the low-weight bit of its left output.
4. Now all the remaining boxes can be successively resolved since they have at least two determined inputs/outputs. The complexity of this attack is then 2^5 while it was 2^6 when all inputs and outputs were considered as elements of $\mathbf{F}_4$.

This attack is successful because the structure of the multipermutation π_2 enables us to deduce some information from the knowledge of only 2 bits amongst all inputs and outputs of $B_{1,0}'$ at the beginning of step 2. This results from a particular property of π_2 which can be expressed in terms of correlation-immune functions.

Proposition 20. *A (r,n) multipermutation π over $\mathbf{F}_{2^m}$ is equivalent to a Boolean function $f_\pi : \mathbf{F}_2^{m(r+n)} \to \mathbf{F}_2$ r-th order correlation-immune with respect to $\mathbf{F}_{2^m}$ such that $|f_\pi^{-1}(1)| = 2^{mr}$. Moreover the nonlinearity order d of f_π verifies:*

$$mn - 1 + \max_{i,j} degree(\pi_i^{(j)}) \leq d \leq m(r+n) - r$$

where $\pi = (\pi_1, \ldots, \pi_n)$ is considered as a function from $\mathbf{F}_2^{mr}$ to $\mathbf{F}_2^{mn}$ and $\pi_i^{(j)}$ is the j-th binary component of π_i.

Proof. A (r,n) multipermutation over $\mathbf{F}_{2^m}$ is equivalent to an orthogonal array of size 2^{mr}, strength r, with $r+n$ constraints over $\mathbf{F}_{2^m}$. This orthogonal array consists of all the $(r+n)$-tuples $(x, \pi(x))$. By Proposition 2 this orthogonal array can be viewed as the truth table of function $f_\pi : \mathbf{F}_2^{m(r+n)} \to \mathbf{F}_2$, which is r-th order correlation-immune with respect to $\mathbf{F}_{2^m}$. By Theorem 17 its normal form is a polynomial of $\mathbf{F}_2[x_i^{(j)}, 1 \leq i \leq r+n, 0 \leq j \leq m-1]$ of degree d and $d + r \leq m(n+r)$. By definition f_π can be expressed by:

$$f_\pi(x) = \prod_{1 \leq i \leq n} \prod_{0 \leq j \leq m-1} [\pi_i^{(j)}(x_1^{(0)}, \ldots, x_r^{(m-1)}) - x_{r+i}^{(j)} - 1]$$

We then deduce the expected inequality.

The complexity of the previous cryptanalysis method, which consists in considering a (r,n) multipermutation π over $\mathbf{F}_{2^m}$ as a function over $\mathbf{F}_2$, can then be deduced from the correlation-immunity order, t, of the Boolean function f_π with respect to $\mathbf{F}_2$. This order has indeed the following cryptographic significance: the knowledge of any $t-1$ bits of inputs and outputs of the box does not permit to determine any of the other bits. For example the previous attack on $g_{2,1}$ results from the fact that the binary correlation-immunity order f_{π_2} is only 2. This attack could be avoided if all multipermutations in the $g_{2,1}$ network were given by the function $f_\pi : \mathbf{F}_4^4 \to \mathbf{F}_2$ whose truth table consists of all codewords of the [4,2,3]-extended Reed-Solomon code over $\mathbf{F}_4$: f_π is indeed a correlation-immune function of order 2 with respect to $\mathbf{F}_4$ and of order 3 with respect to $\mathbf{F}_2$.

Applying Theorem 7 to f_π gives an upper bound on its binary correlation-immunity order depending on its nonlinearity order.

Theorem 21. *Let π be an (r,n) multipermutation over $\mathbf{F}_{2^m}$ and $f_\pi : \mathbf{F}_2^{m(r+n)} \to \mathbf{F}_2$ be the associated Boolean function with nonlinearity order d. Then its binary correlation-immune order t satisfies $r \leq t \leq m(r+n) - d - 1$. In particular we have:*

$$r \leq t \leq mr - \max_{i,j} degree(\pi_i^{(j)})$$

Proof. The binary correlation-immunity order t is obviously greater than r. The second part of the inequality directly comes from Theorem 7 and the associated remark since $|f_\pi^{-1}(1)| = 2^{mr}$, $|f_\pi^{-1}(0)| = 2^{m(n+r)} - 2^{mr}$ and $t < mr$ (this results from the Bush bound [2] which proves the non-existence of binary orthogonal arrays of size 2^{mr}, strength mr with $n(r+m)$ constraints provided $mr > 1$).

This theorem shows that it is not suitable to use multipermutations over $\mathbf{F}_{2^m}$ whose components, considered as Boolean functions, have a high degree.

Example 3.
Let $\pi : \mathbf{F}_8^2 \rightarrow \mathbf{F}_8^2$
$(x; y) \mapsto ((x^3 + y^3)^3; (x^3 + R(y^3) + (y^3 \wedge \alpha))^3)$
where α is a root of $X^3 + X + 1$, R denotes the circular rotation to the right, $+$ is the bitwise XOR and $\wedge$ the bitwise AND.
This function is then a (2,2)-multipermutation over $\mathbf{F}_8$ (see Theorem 4 in [10]). The function $\pi_1^{(0)} : \mathbf{F}_2^6 \rightarrow \mathbf{F}_2$ corresponding to the low-weight component of $\pi_1(x; y) = (x^3+y^3)^3$ has degree 4 as proved in Example 2. The previous theorem gives $2 \leq t \leq 6 - 4$. It follows that the binary correlation-immunity order of f_π is minimal. The use of this multipermutation in a cryptographic primitive is therefore not secure.

References

1. L. Brynielsson. On the linear complexity of combined shift register sequences. In F. Pichler, editor, *Advances in Cryptology - EUROCRYPT '85*, number 219 in Lecture Notes in Computer Science, pages 156–160. Springer-Verlag, 1986.
2. K.A. Bush. Orthogonal arrays of index unity. *Ann. Math. Stat.*, 23:426–434, 1952.
3. P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In J. Feigenbaum, editor, *Advances in Cryptology - CRYPTO'91*, number 576 in Lecture Notes in Computer Science, pages 86–100. Springer-Verlag, 1992.
4. K. Gopalakrishnan and D.R. Stinson. Three characterizations of non-binary correlation-immune and resilient functions. *Designs, Codes and Cryptography*, 5:241–251, 1995.
5. T. Herlestam. On functions of linear shift register sequences. In F. Pichler, editor, *Advances in Cryptology - EUROCRYPT '85*, number 219 in Lecture Notes in Computer Science, pages 119–129. Springer-Verlag, 1986.
6. U.M. Maurer and J.L. Massey. Perfect local randomness in pseudo-random sequences. In G. Brassard, editor, *Advances in Cryptology - CRYPTO'89*, number 435 in Lecture Notes in Computer Science, pages 100–112. Springer-Verlag, 1990.
7. C.R. Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *J. Roy. Statist.*, 9:128–139, 1947.
8. R.A. Rueppel. *Analysis and Design of stream ciphers*. Springer-Verlag, 1986.
9. R.A. Rueppel and O.J. Staffelbach. Products of linear recurring sequences with maximum complexity. *IEEE Trans. Inform. Theory*, IT-33(1):124–131, 1987.
10. C.-P. Schnorr and S. Vaudenay. Black box cryptanalysis of hash networks based on multipermutations. In A. De Santis, editor, *Advances in Cryptology - EUROCRYPT'94*, number 950 in Lecture Notes in Computer Science, pages 47–57. Springer-Verlag, 1995.
11. C.P. Schnorr and S. Vaudenay. Parallel FFT-Hashing. In *Fast Software Encryption*, number 809 in Lecture Notes in Computer Science, pages 149–156. Springer-Verlag, 1994.
12. T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Trans. Inform. Theory*, IT-30(5):776–780, 1984.

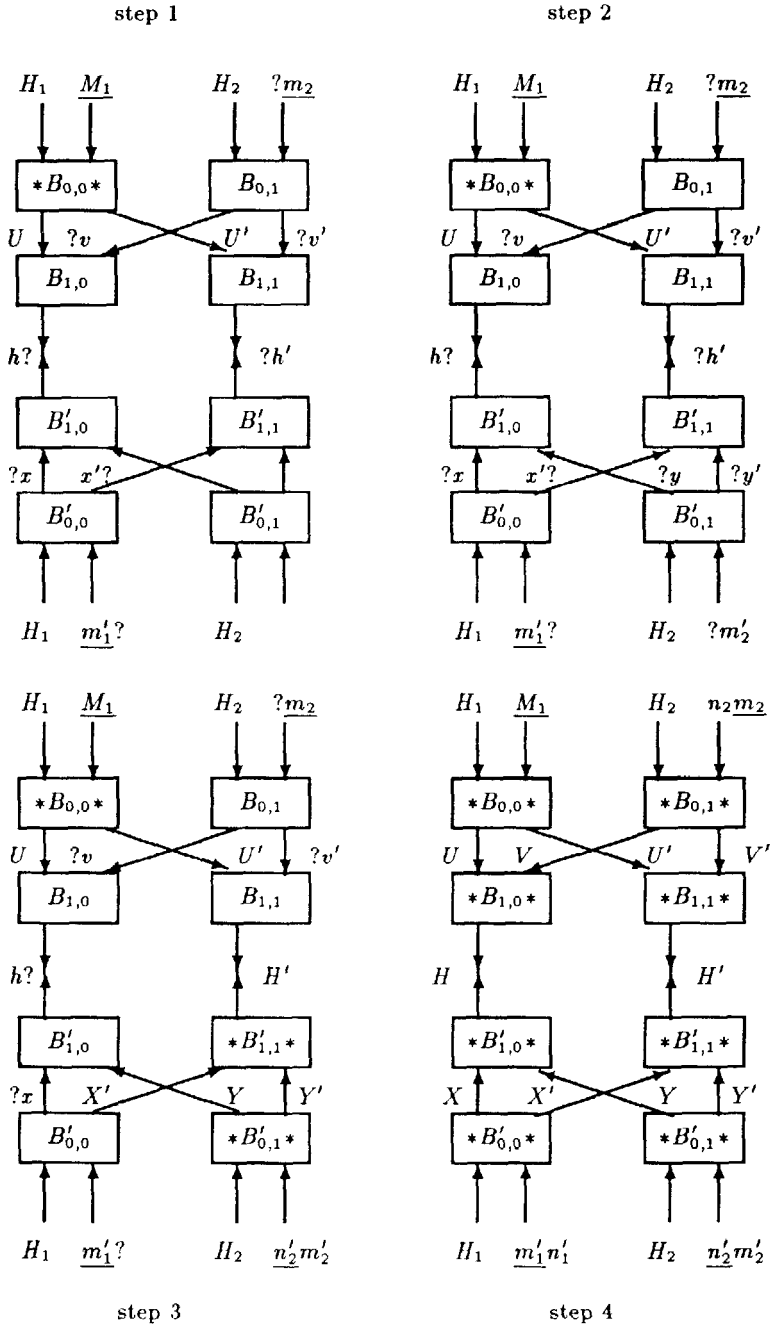

Fig. 3. Resolution of the collision network for $g_{2,1}$