IOP Conf. Series: Journal of Physics: Conf. Series **855** (2017) 012032 doi:10.1088/1742-6596/855/1/012032

Error-correcting pairs for a public-key cryptosystem

Ruud Pellikaan

Dept. of Mathematics and Computing Science, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

E-mail: g.r.pellikaan@tue.nls

Irene Márquez-Corbella

Dept. of Mathematics, Statistics and O. Research, University of La Laguna, Spain

E-mail: irene.marquez.corbella@ull.es

Abstract. Code-based Cryptography (CBC) is a powerful and promising alternative for quantum resistant cryptography. Indeed, together with lattice-based cryptography, multivariate cryptography and hash-based cryptography are the principal available techniques for post-quantum cryptography. CBC was first introduced by McEliece where he designed one of the most efficient Public-Key encryption schemes with exceptionally strong security guarantees and other desirable properties that still resist to attacks based on Quantum Fourier Transform and Amplitude Amplification.

The original proposal, which remains unbroken, was based on binary Goppa codes. Later, several families of codes have been proposed in order to reduce the key size. Some of these alternatives have already been broken.

One of the main requirements of a code-based cryptosystem is having high performance t-bounded decoding algorithms which is achieved in the case the code has a t-error-correcting pair (ECP). Indeed, those McEliece schemes that use GRS codes, BCH, Goppa and algebraic geometry codes are in fact using an error-correcting pair as a secret key. That is, the security of these Public-Key Cryptosystems is not only based on the inherent intractability of bounded distance decoding but also on the assumption that it is difficult to retrieve efficiently an error-correcting pair.

In this paper, the class of codes with a t-ECP is proposed for the McEliece cryptosystem. Moreover, we study the hardness of distinguishing arbitrary codes from those having a t-error correcting pair.

1. Introduction

In 1978 [17] McEliece presented the first PKC system based on the theory of error-correcting codes. In 1986 Niederreiter [19] presented a dual version of McEliece cryptosystem which is equivalent in terms of security. Their main advantages are its fast encryption and decryption schemes. It is an interesting candidate for post-quantum cryptography.

2. Code-based cryptography

A *linear code* C is a subspace of $\mathbb{F}_q^n$. The *weight* of $\mathbf{x} \in \mathbb{F}_q^n$ is the number of nonzero entries of $\mathbf{x}$ and is denoted by $\mathrm{wt}(\mathbf{x})$. The *(Hamming) distance* between $\mathbf{x}$, $\mathbf{y} \in \mathbb{F}_q$ is the number of entries where $\mathbf{x}$ and $\mathbf{y}$ differ and is denoted by $d(\mathbf{x}, \mathbf{y})$. The *minimum distance* of C is the minimal

value of $d(\mathbf{x}, \mathbf{y})$ where $\mathbf{x},\ \mathbf{y} \in C$ and $\mathbf{x} \neq \mathbf{y}$. Since C is linear it is equal to the minimum weight of C, that is the minimal value of $\mathrm{wt}(\mathbf{x})$ where $\mathbf{x} \in C$ and $\mathbf{x} \neq \mathbf{0}$.

The *parameters* of the code are denoted by $[n, k, d]$, where n is its *length*, k its *dimension* and d its *minimum distance*. The *(information) rate* of C is defined by $R = k/n$.

Let C be an $\mathbb{F}_q$-linear code of length n and dimension k. A *generator matrix* G of C is a $k \times n$ matrix with entries in $\mathbb{F}_q$ such that its rows are a basis of C. A *parity check matrix* H is $(n-k) \times n$ matrix with entries in $\mathbb{F}_q$ such that $\mathbf{c}H^T = \mathbf{0}$ if and only if $\mathbf{c} \in C$.

The problem of *minimum distance decoding* has as input $(G, \mathbf{y})$, where G is a generator matrix of a code C over $\mathbb{F}_q$ of parameters $[n, k, d]$ and $\mathbf{y} \in \mathbb{F}_q^n$ is a received word. The output is a codeword $\mathbf{c} \in C$ of minimal distance to $\mathbf{y}$. One can phrase the problem equivalently in terms of a parity check matrix H of the code. Then the input is $(H, \mathbf{s})$, where $\mathbf{s} \in \mathbb{F}_q^{n-k}$. The output is an $\mathbf{e} \in \mathbb{F}_q^n$ of minimal weight such that $\mathbf{e}H^T = \mathbf{s}$. The relation of the two versions is given by $\mathbf{s} = \mathbf{y}H^T$ the *syndrome* and $\mathbf{e} = \mathbf{y} - \mathbf{c}$ the error vector of the received word $\mathbf{y}$.

The security of code-based cryptosystems is based on the hardness of decoding up to half the minimum distance. The minimum distance decoding problem was shown by Berlekamp-McEliece-Van Tilborg [2] to be NP-hard. The status of the hardness of decoding up to half the minimum distance is an open problem. McEliece proposed to use binary Goppa codes for his PKC system.

In the McEliece PKC system a collection $\mathcal{K}$ of generator $k \times n$ matrices is chosen for which an efficient decoding algorithm is available that corrects all patterns of t errors. The *encryption map*

$$E_G : \ \mathcal{P} \ \rightarrow \ \mathcal{C}$$

for a given key $G \in \mathcal{K}$ is defined by $E_G(\mathbf{m}, \mathbf{e}) = \mathbf{m}G + \mathbf{e}$. An *adversary* A is a map from $\mathcal{C} \times \mathcal{K}$ to $\mathcal{P}$. This adversary is successful for $(x, G) \in \Omega$ if $A(E_G(x), G) = x$.

Let $\mathcal{C}$ be a class of codes such that every code C in $\mathcal{C}$ has an efficient decoding algorithm correcting all patterns of t errors. Let $G \in \mathbb{F}_q^{k \times n}$ be a generator matrix of C. In order to mask the origin of G, take a $k \times k$ invertible matrix S over $\mathbb{F}_q$ and an $n \times n$ permutation or monomial matrix Π. Then for the McEliece PKC the matrices G, S and Π are kept secret while $G' = SG\Pi$ is made public. Furthermore the (trapdoor) one-way function of this cryptosystem is usually presented as follows:

$$\mathbf{x} = (\mathbf{m}, \mathbf{e}) \mapsto \mathbf{y} = \mathbf{m}G' + \mathbf{e},$$

where $\mathbf{m} \in \mathbb{F}_q^k$ is the plaintext and $\mathbf{e} \in \mathbb{F}_q^n$ is a random error vector with Hamming weight at most t.

3. Error-correcting pairs

From now on the dimension of a linear code C will be denoted by $k(C)$ and its minimum distance by $d(C)$. Given two elements $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{F}_q^n$, the *star product* is defined by coordinatewise multiplication, that is $\mathbf{a} * \mathbf{b} = (a_1 b_1, \ldots, a_n b_n)$ while the *standard inner multiplication* is defined by $\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i$.

Let A, B and C be subspaces of $\mathbb{F}_q^n$. Then $A * B$ is the subspace generated by $\{\mathbf{a} * \mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B\}$. And $C^\perp = \{\mathbf{x} | \mathbf{x} \cdot \mathbf{c} = 0 \text{ for all } \mathbf{c} \in C\}$ is the *dual* code of C. Furthermore $A \perp B$ means $\mathbf{a} \cdot \mathbf{b} = 0$ for all $\mathbf{a} \in A$ and $\mathbf{b} \in B$.

Definition 3.1. Let C be a linear code in $\mathbb{F}_q^n$. The pair (A, B) of linear codes over $\mathbb{F}_{q^m}$ of length n is called a *t-error-correcting pair* (ECP) over $\mathbb{F}_{q^m}$ for C if the following properties hold:

E.1 $(A * B) \perp C$,

E.2 $k(A) > t$,

E.3 $d(B^\perp) > t$,

E.4 $d(A) + d(C) > n$.

Remark 3.2. In the above definition A and B are $\mathbb{F}_{q^m}$-linear codes and C is an $\mathbb{F}_q$-linear code. So by $k(A)$ the dimension of A over $\mathbb{F}_{q^m}$ is meant. And $d(B^\perp)$, $d(A)$ and $d(C)$ mean the minimum distances of $B^\perp$ and A over $\mathbb{F}_{q^m}$ and of C over $\mathbb{F}_q$.

Remark 3.3. The notion of an error-correcting pair for a linear code was introduced in 1988 by Pellikaan [22] and independently by Kötter in [11, 12] in 1992. It is shown that a linear code in $\mathbb{F}_q^n$ with a t-error-correcting pair has a decoding algorithm which corrects up to t errors with complexity $\mathcal{O}(n^3)$.

Remark 3.4. Note that if (A, B) is a pair of codes that satisfies Conditions E.1, E.2, E.3 and the following two conditions:

E.5 $d(A^\perp) > 1$*, that means* A *is a non-degenerate code,*
E.6 $d(A) + 2t > n$*,*

then $d(C) \geq 2t + 1$ and (A, B) is a t-ECP for C by [23, Corollary 3.4].

In the following we consider eight collections of pairs.

Example 3.5. Let $\mathbf{a}$ be an n-tuple of mutually distinct elements of $\mathbb{F}_q$ and $\mathbf{b}$ be an n-tuple of nonzero elements of $\mathbb{F}_q$. Then the *generalized Reed-Solomon* code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is defined by

$$\mathrm{GRS}_k(\mathbf{a}, \mathbf{b}) = \{(f(a_1)b_1, \ldots, f(a_n)b_n) \mid f(X) \in \mathbb{F}_q[X] \text{ and } \deg(f(X)) < k\}\,.$$

If $k \leq n \leq q$, then $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is an $[n, k, n - k + 1]$ code. Furthermore the dual of a GRS code is again a GRS code, in particular $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})^\perp = \mathrm{GRS}_{n-k}(\mathbf{a}, \mathbf{b}^\perp)$ for some $\mathbf{b}^\perp$ that is explicitly known.

Let $A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{u})$, $B = \mathrm{GRS}_t(\mathbf{a}, \mathbf{v})$ and $C = \mathrm{GRS}_{2t}(\mathbf{a}, \mathbf{u} * \mathbf{v})^\perp$. Then (A, B) is a t-ECP for C. Conversely let $C = \mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$, then $A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{b}^\perp)$ and $B = \mathrm{GRS}_t(\mathbf{a}, \mathbf{1})$ is a t-ECP for C where $t = \left\lfloor \frac{n-k}{2} \right\rfloor$.

So GRS codes are the prime examples of codes that have a t-error-correcting pair. GRS codes are not suited for a coded-based PKC by the attack of Sidelnikov-Shestakov [26].

Example 3.6. Let C be a subcode of a code D that has (A, B) as a t-ECP. Then condition (E.1) holds for (A, B) with respect to D. So $\mathbf{a} * \mathbf{b} \cdot \mathbf{d} = 0$ for all $\mathbf{d}$ in D. Hence $\mathbf{a} * \mathbf{b} \cdot \mathbf{c} = 0$ for all $\mathbf{c}$ in C, since $C \subseteq D$. Conditions (E.2), (E.3) and (E.4) hold. Therefore (A, B) is also a t-ECP for C.

In particular, let C be a subcode of the code $\mathrm{GRS}_{n-2t}(\mathbf{a}, \mathbf{b})$. This GRS code has a t-error-correcting pair by Example 3.5 which is also a t-ECP for C.

The class of subcodes of GRS codes was proposed by Berger and Loidreau in [1] for code-based PKC to resist precisely the Sidelnikov-Shestakov attack. But for certain parameter choices this proposal is also not secure as shown by Márquez et al. [14].

Example 3.7. The Goppa code $\Gamma(L, g(X))$ associated to a Goppa polynomial $g(X)$ of degree r and an n-tuple L of points in $\mathbb{F}_{q^m}$ can be viewed as an alternant code, that is a subfield subcode of a GRS code of codimension r. Therefore such a code has an $\lfloor r/2 \rfloor$-error-correcting pair. If the Goppa polynomial is square free of degree r in an extension of $\mathbb{F}_2$, then the binary Goppa code has an r-ECP, since $\Gamma(L, g(X)) = \Gamma(L, g(X)^2)$.

Goppa codes were proposed by McEliece [17] for his PKC system. Sidelnikov-Shestakov made a claim [26] that their method for GRS codes could be extended to attack Goppa codes as well, but this had never been substantiated by a paper in the public domain. A binary Goppa code using elements in the extension $\mathbb{F}_{2^m}$ and with a square free Goppa polynomial of degree t over $\mathbb{F}_{2^m}$ has parameters $[n, k, d]$ with $n \leq 2^m$, $k \geq n - mt$ and $d \geq 2t + 1$. For these codes efficient decoding algorithms are known that decode all patterns with t errors.

A binary Goppa code with parameters $[1024, 524, 101]$ as proposed by McEliece is no longer secure with nowadays computing power due to recent improvements in the decoding algorithms.

Example 3.8. Algebraic geometry (AG) codes were introduced in 1977 by V.D. Goppa. Recall that GRS codes can be seen as the class of AG codes on the projective line, that is the algebraic curve of genus zero. We refer the interested reader to [24, 25].

Let $\mathcal{X}$ be an algebraic curve defined over $\mathbb{F}_q$ with genus g. By an algebraic curve we mean a curve that is absolutely irreducible, nonsingular and projective. Let $\mathcal{P}$ be an n-tuple of $\mathbb{F}_q$-rational points on $\mathcal{X}$ and let E be a divisor of $\mathcal{X}$ with disjoint support from $\mathcal{P}$ of degree e. Then the algebraic geometry code $C_L(\mathcal{X}, \mathcal{P}, E)$ is the image of the Riemann-Roch space $L(E)$ of rational functions with prescribed behavior of zeros and poles at E under the evaluation map $\mathrm{ev}_{\mathcal{P}}$. If $e < n$, then the dimension of the code $C_L(\mathcal{X}, \mathcal{P}, E)$ is at least $e + 1 - g$ and its minimum distance is at least $n - e$. If $e > 2g - 2$, then its dimension is $e + 1 - g$. The dual code $C_L(\mathcal{X}, \mathcal{P}, E)^\perp$ is again AG. If $e > 2g - 2$, then the dimension of the code $C_L(\mathcal{X}, \mathcal{P}, E)^\perp$ is at least $n - e - 1 + g$ and its minimum distance is at least $d^* = e - 2g + 2$. If $e < n$, then its dimension is $n - e - 1 + g$.

If $A = C_L(\mathcal{X}, \mathcal{P}, E)$ and $B = C_L(\mathcal{X}, \mathcal{P}, F)$, then $A * B \subseteq C_L(\mathcal{X}, \mathcal{P}, E + F)$. So there are abundant ways to construct error-correcting pairs of an AG code. An AG code on a curve of genus g with designed minimum distance d^* has a t-ECP over $\mathbb{F}_q$ with $t = \lfloor (d^* - 1 - g)/2 \rfloor$ by [21, Theorem 1] and [22, Theorem 3.3]. If m is sufficiently large, then there exists a t-ECP over $\mathbb{F}_{q^m}$ with $t = \lfloor (d^* - 1)/2 \rfloor$ by [23, Proposition 4.2].

Algebraic geometry codes were proposed by Niederreiter [19] and Janwa-Moreno [10] for code-based PKC systems. This system was broken for low genus zero [26], one and two [9, 18]. For arbitrary genus it was shown by Márquez et al. [13, 15] that these codes are not secure for rates R in the intervals $[\gamma, \frac{1}{2} - \gamma]$, $[\frac{1}{2} + \gamma, 1 - \gamma]$, $[\frac{1}{2} - \gamma, 1 - 3\gamma]$ and $[3\gamma, \frac{1}{2} + \gamma]$, where $R = k/n$ is the information rate and $\gamma = g/n$ the relative genus. Recently Couvreur et al. [6] showed that it is not necessary to retrieve the triple $(\mathcal{X}, \mathcal{P}, E)$ and the Riemann-Roch space $L(E)$ but that one can stay in the realm of $\mathbb{F}_q^n$ and its subspaces in order to find an error-correcting pair.

Example 3.9. Geometric Goppa codes are subfield subcodes of algebraic geometry codes generalizing the classical Goppa codes that are subfield subcodes of GRS codes. Geometric Goppa codes were proposed by Janwa-Moreno [10]. Couvreur et al. [5] showed that certain geometric Goppa codes are not secure for a PKC system.

Example 3.10. Let (A, B) be a pair of codes with parameters $[n, t+1, n-t]$ and $[n, t, n-t+1]$, respectively, and $C = (A * B)^\perp$, then the minimum distance of C is at least $2t + 1$ and (A, B) is a t-error-correcting pair for C by [23, Corollary 3.4]. The dimension of $A * B$ is at most $t(t+1)$. So the dimension of C is at least $n - t(t + 1)$. In Appendix A it is shown that this is almost always equal to $n - t(t + 1)$ for random choices of A and B.

If q is considerably larger than n, then a random linear code is MDS with very high probability. So taking random codes A and B of length n and dimensions $t + 1$ and t, respectively, this gives a very large class of codes for the McEliece PKC. However with large field the key size becomes larger and recall that the main obstacle for coded-based cryptosystems was the key size.

4. The ECP one-way function

Let $\mathcal{P}(n, t, q)$ be the collection of pairs (A, B) such that there exist a positive integer m and a pair (A, B) of $\mathbb{F}_{q^m}$-linear codes of length n, that satisfy Conditions E.2, E.3, E.5 and E.6. Let C be the $\mathbb{F}_q$-linear code of length n that is the subfield subcode that has all elements of $A * B$ as parity checks. So

$$C = \mathbb{F}_q^n \cap (A * B)^\perp.$$

Then the minimum distance of C is at least $2t + 1$ and (A, B) is a t-ECP for C as was noted in Remark 3.4. Let $\mathcal{F}(n, t, q)$ be the collection of $\mathbb{F}_q$-linear codes of length n and minimum distance

$d \geq 2t+1$. Consider the following map

$$\begin{array}{rccc} \varphi_{(n,t,q)}: & \mathcal{P}(n,t,q) & \longrightarrow & \mathcal{F}(n,t,q) \\ & (A,B) & \longmapsto & C. \end{array}$$

The question is whether this map is a one-way function.

Let U and V be generator matrices of the codes A and B, with rows denoted by $\mathbf{u}_i$ and $\mathbf{v}_i$, respectively. Let $U*V$ be the matrix with the rows $\mathbf{u}_i * \mathbf{v}_j$ ordered lexicographically. Let $(U*V)(q^l)$ be the matrix with entries the q^l-power of the entries of $U*V$. Let W be the reduced row echelon form (with the zero rows deleted) of the matrix with rows all the rows of $(U*V)(q^l)$ for $l = 0, 1 \ldots, m-1$. Then W has entries in $\mathbb{F}_q$ and is a parity check matrix of C. In this way

$$(U,V) \longmapsto \quad W$$

is an implementation of the map $\varphi_{(n,t,q)}$.

If the map $\varphi_{(n,t,q)}$ is indeed difficult to invert, then we will call it the *ECP one-way function* and the code C with parity check matrix W might be used as a public-key in a coding based PKC. Otherwise it would mean that the PKC based on codes that can be decoded by error-correcting pairs is not secure.

Remark 4.1. Note that $\mathbf{u}\Pi * \mathbf{v}\Pi = (\mathbf{u} * \mathbf{v})\Pi$ for every permutation or monomial matrix Π. Thus, if (A,B) is a t-ECP for C, then $(A\Pi, B\Pi)$ is a t-ECP for $C\Pi$.

Furthermore, if S and T are invertible matrices of the correct sizes to be multiplied on the left of the matrices U and V, respectively, then $U*V$ generates the same code as $(SU)*(TV)$ since $(SU)*\mathbf{v} = S(U*\mathbf{v})$ and $\mathbf{u}*(TV) = T(\mathbf{u}*V)$ for all vectors $\mathbf{u}$ and $\mathbf{v}$. Therefore the usual masking SHP of a parity check matrix H by means of an invertible matrix S and a permutation matrix P is already incorporated in the choice of the pair of generator matrices (U,V).

5. Distinguishing a code with an ECP

Let $\mathcal{K}$ be a collection of generator matrices of codes that have a t-error-correcting pair and that is used for a coded-based PKC system. In this section we address assumption A.2 whether we can distinguish arbitrary codes from those coming from $\mathcal{K}$.

Let C be a k dimensional subspace of $\mathbb{F}_q^n$ with basis $\mathbf{g}_1, \ldots, \mathbf{g}_k$ which represents the rows of the generator matrix $G \in \mathbb{F}_q^{k\times n}$. We denote by $S^2(C)$ the *second symmetric power* of C, or equivalently the *symmetrized tensor product* of C with itself. If $\mathbf{x}_i = \mathbf{g}_i$, then $S^2(C)$ has basis $\{\mathbf{x}_i\mathbf{x}_j \mid 1 \leq i \leq j \leq n\}$ and dimension $\binom{k+1}{2}$. Furthermore we denote $C*C$ by $C^{(2)}$ the *square* of C, that is the linear subspace in $\mathbb{F}_q^n$ generated by $\{\mathbf{a}*\mathbf{b} | \mathbf{a},\mathbf{b} \in C\}$. See [3, §4 Definition 6] and [4, 14]. Now, following the same scheme as in [13], we consider the linear map

$$\sigma: \quad S^2(C) \quad \longrightarrow \quad C^{(2)},$$

where the element $\mathbf{x}_i\mathbf{x}_j$ is mapped to $\mathbf{g}_i * \mathbf{g}_j$. The kernel of this map will be denoted by $K^2(C)$. Then $K^2(C)$ is the solution space of the following set of equations:

$$\sum_{1\leq i\leq i'\leq k} g_{ij}g_{i'j}\mathbf{X}_{ii'} = 0, \quad 1 \leq j \leq n.$$

There is no loss of generality in assuming G to be systematic at the first k position, making a suitable permutation of columns and applying Gaussian elimination, if necessary. Then $G = \begin{pmatrix} I_k & P \end{pmatrix}$ where I_k is the $k \times k$ identity matrix and P is an $k \times (n-k)$ matrix formed by the last $n-k$ columns of G. Now $H = \begin{pmatrix} P^T & -I_{n-k} \end{pmatrix}$ is a parity check matrix of C, or equivalently H is a generator matrix of the $[n, n-k]$ code $D = C^{\perp}$.

In [8, III] and [20, Ch. 10] a system $\mathcal{L}_P$ associated to the matrix P of k linear equations involving the $\binom{n-k}{2}$ variables Z_{jl}, with $k+1 \leq j < l \leq n$, is defined as

$$\mathcal{L}_P = \left\{ \sum_{j=k+1}^{n-1} \sum_{j'>j}^{n} p_{ij} p_{ij'} \mathbf{Z}_{jj'} = 0 \mid 1 \leq i \leq k. \right\}$$

This system differs from the system of equations obtained for the kernel $K^2(C)$ in interchanging indices i and j and the strict inequality $j < j'$ in the summation, instead of $i \leq i'$. Denote the kernel of $\mathcal{L}_P$, that is the space of all solutions of $\mathcal{L}_P$, by $K(\mathcal{L}_P)$.

Proposition 5.1.

$$\dim K(\mathcal{L}_P) = \dim K^2(D)$$

Proof. Let M be the $\binom{k+1}{2} \times n$ matrix with entries $\left(g_{ij} g_{i'j}\right)_{\substack{1 \leq i \leq i' \leq k \\ 1 \leq j \leq n}}$. Then a basis of $K^2(C)$ can be read of directly as the kernel of M. Note also that the dimension of $C^{(2)}$ is equal to the rank of M. Furthermore, since $C^{(2)}$ is the image of the linear map σ, by the first isomorphism theorem we get:

$$\dim K^2(C) + \dim C^{(2)} = \dim S^2(C) = \binom{k+1}{2}.$$

Let $\mathbf{h}_i$ be the i-th row of the parity check matrix H, $\mathbf{e}_i$ be the i-th vector in the canonical basis of $\mathbb{F}_q^{n-k}$ and $\mathbf{q}_i$ be the i-th row of the matrix P^T. Then $q_{ij} = p_{j,i+k}$ and $\mathbf{h}_i = (\mathbf{q}_i | - \mathbf{e}_i)$. Therefore

$$\mathbf{h}_j * \mathbf{h}_{j'} = \begin{cases} \left(\mathbf{q}_j * \mathbf{q}_j \mid \mathbf{e}_j \right) & \text{if } j = j', \\ \left(\mathbf{q}_j * \mathbf{q}_{j'} \mid \mathbf{0} \right) & \text{if } j < j'. \end{cases}$$

Let M_1 be the $k \times \binom{n-k}{2}$ matrix with entries $\left(p_{ij} p_{ij'}\right)_{\substack{1 \leq i \leq k \\ k < j < j' \leq n}}$, then

$$\dim K(\mathcal{L}_P) = \binom{n-k}{2} - \operatorname{rank}(M_1)$$

Now let M_2 be the $\binom{n-k+1}{2} \times n$ matrix with entries $\left(h_{ij} h_{i'j}\right)_{\substack{1 \leq i \leq i' \leq n-k \\ 1 \leq j \leq n}}$. Then

$$\dim D^{(2)} = \operatorname{rank}(M_2) = n - k + \operatorname{rank}(M_1)$$

Therefore

$$\begin{aligned} \dim K(\mathcal{L}_P) &= \binom{n-k}{2} - \operatorname{rank}(M_1) \\ &= \binom{n-k}{2} + n - k - \dim D^{(2)} \\ &= \dim K^2(D) \end{aligned}$$

□

The dual statement of Proposition 5.1 gives: $\dim K(\mathcal{L}_{P^T}) = \dim K^2(C)$.
For every $[n, k]$ code C over $\mathbb{F}_q$ the following inequality holds:

$$\dim C^{(2)} \leq \min\{n, \tbinom{k+1}{2}\}.$$

However if the entries of the matrix P are taken independently and identically distributed, then the inequality holds with equality with high probability what is actually proved in the next proposition.

Proposition 5.2. *Let C be an $[n,k]$ code with $n > \binom{k+1}{2}$ chosen at random. Then*

$$\Pr\left(\dim(C^{(2)}_{random}) = \binom{k+1}{2}\right) = o(1)$$

Proof. Let C be a linear code with parameters $[n,k]$ over $\mathbb{F}_q$ with $n > \binom{k+1}{2}$.

We have seen in the proof of Proposition 5.1, with the role of C and $D = C^\perp$ interchanged that the linear system $\mathcal{L}_{P^T}$ associated with C consists of $n-k$ linear equations and $\binom{k}{2}$ unknowns. In case $n-k > \binom{k}{2}$ or equivalently $n > \binom{k+1}{2}$ Faugère et al. [8] proved that the dimension of the solution space of $\mathcal{L}_{P^T}$ is 0 with high probability. Therefore under the same hypothesis we have that the dimension of $C^{(2)}_{\text{random}}$ is $\binom{k+1}{2}$ with high probability. □

Remark 5.1. See [4, Theorem 2.3] for an improved version of Proposition 5.2.

Example 5.2. Let C be a GRS code with parameters $[n,k]$, take for instance $C = \mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ where $\mathbf{a}$ is an n-tuple of mutually distinct elements of $\mathbb{F}_q$ and $\mathbf{b}$ is an n-tuple of nonzero elements of $\mathbb{F}_q$. Then $C^{(2)}$ is the code $\mathrm{GRS}_{2k-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$ if $2k-1 \leq n$ and $\mathbb{F}_q^n$ otherwise. Hence $\dim C^{(2)} = \min\{2k-1, n\}$. Therefore

$$\dim K^2(C) = \binom{k+1}{2} - (2k-1) = \binom{k-1}{2} \text{ if } 2k-1 \leq n.$$

Example 5.3. Let C be a k-dimensional subcode of the code $\mathrm{GRS}_l(\mathbf{a},\mathbf{b})$. Then $C^{(2)}$ is a subcode of the code $\mathrm{GRS}_{2l-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$, if $2l-1 \leq n$. Thus

$$\dim C^{(2)} \leq \min\{2l-1, n\}.$$

Moreover if $4l-3k-1 < q$ and $2l-1 \leq \binom{k+1}{2}$, then it was shown in [14] that $C^{(2)}$ is equal to $\mathrm{GRS}_{2l-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$ with high probability so, under this hypothesis,

$$\Pr\left(\dim C^{(2)} = 2l-1\right) = 1 - o(1).$$

The dual code $D = C^\perp$ contains the code $\mathrm{GRS}_l(\mathbf{a},\mathbf{b})^\perp = \mathrm{GRS}_{n-l}(\mathbf{a},\mathbf{b}^\perp)$. That is, $D^{(2)}$ contains the square of $\mathrm{GRS}_{n-l}(\mathbf{a},\mathbf{b}^\perp)$ which is equal to $\mathrm{GRS}_{2n-2l-1}(\mathbf{a},\mathbf{b}^\perp*\mathbf{b}^\perp)$ if $2n-2l-1 \leq n$, or equivalently if $n \leq 2l+1$. Recall that the star product of the rows of a generator matrix of any linear code gives a generating set for its square code, that is the square of any $[n,s]$ linear code is generated by $\binom{s+1}{2}$ elements. In particular $D^{(2)}$ is generated by $\binom{n-k+1}{2}$ elements but since $\mathrm{GRS}_{2n-2l-1}(\mathbf{a},\mathbf{b}^\perp*\mathbf{b}^\perp) \subseteq D^{(2)}$ there are at least $\binom{n-l+1}{2} - (2n-2l+1)$ dependent elements of this generating set. Thus

$$\dim D^{(2)} \leq \binom{n-k+1}{2} - \binom{n-l+1}{2} + 2n - 2l - 1 = \binom{n-k+1}{2} - \binom{n-l-1}{2}.$$

Example 5.4. The problem of distinguishing Goppa, alternant and random codes from each other was studied by Faugère et al. in [8]. Their experimental results give rise to a conjecture on the dimension of $K(\mathcal{L}_P)$ for Goppa and alternant codes of high rate.

Example 5.5. Let $C = \mathcal{C}_L(\mathcal{X},\mathcal{P},E)$ where $\mathcal{X}$ is an algebraic curve over $\mathbb{F}_q$ of genus g, $\mathcal{P}$ is an n-tuple of mutually distinct $\mathbb{F}_q$-rational points of $\mathcal{X}$ and E is a divisor of $\mathcal{X}$ with disjoint support from $\mathcal{P}$ of degree e. Then $C^{(2)} \subseteq \mathcal{C}_L(\mathcal{X},\mathcal{P},2E)$.
Assume moreover that $2g-2 < e < n/2$. Then C has dimension $k = e+1-g$ and $C_L(\mathcal{X},\mathcal{P},2E)$

has dimension $2e+1-g = k+e$. Hence $\dim C^{(2)} \leq k+e$.
Let G be a generator matrix of an algebraic geometry code C. Take the columns of G as homogeneous coordinates of points in $\mathbb{P}^{e-g}$, this gives a projective system $\mathcal{Q} = (Q_1, \ldots, Q_n)$ of points in the projective space $\mathbb{P}^{e-g}(\mathbb{F}_q)$. Since $e > 2g$ there exists an embedding of the curve $\mathcal{X}$ in $\mathbb{P}^{e-g}$ of degree e

$$\begin{array}{rccc} \varphi_E: & \mathcal{X} & \longrightarrow & \mathbb{P}^{e-g} \\ & P & \longmapsto & \varphi_E(P) = (f_0(P), \ldots, f_{e-g}(P)) \end{array}$$

where $\{f_0, \ldots, f_{e-g}\}$ is a basis of $L(E)$ such that $\mathcal{Q} = \varphi_E(\mathcal{P})$ lies on the curve $\mathcal{Y} = \varphi_E(\mathcal{X})$. The space $I_2(\mathcal{Q})$ of quadratic polynomials that vanish on $\mathcal{Q}$ can be identified with $K^2(C)$. Furthermore if $2g+2 \leq e < \frac{1}{2}n$, then $I_2(\mathcal{Y}) = I_2(Q)$ and $I(\mathcal{Y})$, the vanishing ideal of $\mathcal{Y}$, is generated by $I_2(Q)$. Now

$$\dim K^2(C) = \binom{k+1}{2} - \dim C^{(2)} \geq \binom{k}{2} - e.$$

Therefore $\mathcal{Y}$ is given as the intersection of at least $\binom{k}{2} - e$ quadrics in $\mathbb{P}^{e-g}$. For more details we refer the reader to [13, 15].

Example 5.6. Let $t(t+1) < n$. Let (A, B) be a pair of random codes of dimension $t+1$ and t, respectively. Take $C = (A * B)^{\perp}$ as in Example 3.10. Let $D = C^{\perp} = A * B$. Then $D^{(2)} = A^{(2)} * B^{(2)}$. Hence

$$\dim D^{(2)} \leq \binom{t+2}{2}\binom{t+1}{2}$$

which is about half the expected value $\binom{t(t+1)}{2}$ in case $\binom{t(t+1)}{2} < n$ by Proposition 5.2, since D has dimension $t(t+1)$ with high probability by Appendix A.

6. Acknowledgement

An earlier version of this work was presented for the first time by the second author at the *Code-Based Cryptography Workshop*, May 2012 at the Technical University of Denmark, Lyngby and posted at arXiv [16] and furthermore at the conferences Applications of Computer Algebra 2013 and 2014 at Malaga and Fordham, respectively, and finally at the 3rd IndoMS International Conference on Mathematics and Its Applications, Depok in 2015.

References

[1] Berger, T., Loidreau, P.: How to mask the structure of codes for a cryptographic use. Designs, Codes and Cryptography 35, 63–79 (2005)
[2] Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems. IEEE Transactions on Information 24, 384–386 (1978)
[3] Cascudo, I., Chen, H., Cramer, R., Xing, X.: Asymptotically good ideal linear secret sharing with strong multiplication overy any fixed finite field. In: Halevi, S. (ed.) Advances in Cryptology - CRYPTO 2009, Lecture Notes in Computer Science. vol. 5677, pp. 466–486. Springer, Berlin (2009)
[4] Cascudo, I., Cramer, R., Mirandola, D., Zémor, G.: Squares of random linear codes. IEEE Trans. Inform. Theory 61(3), 1159–1173 (2015)
[5] Couvreur, A., Márquez-Corbella, I., Pellikaan, R.: Cryptanalysis of public-key cryptosystems that use subcodes of algebraic geometry codes. In: 4ICMCTA, Coding Theory and Application, CIM Series in Mathematical Sciences 3, pp. 133–140 (2014)
[6] Couvreur, A., Márquez-Corbella, I., Pellikaan, R.: A polynomial time attack against algebraic geometry code based public key cryptosystems. In: IEEE International Symposion on Information Theory ISIT 2014, p. 1446 (2014)
[7] Faugère, J., Gauthier-Umana, V., Otmani, A., Perret, L., Tillich, J.: A distinguisher for high rate McEliece cryptosystems. In: Proceedings IEEE Information Theory Workshop 2011. Paraty, Brazil (October 16-20, 2011)

[8] Faugère, J.C., Gauthier-Umaña, V., Otmani, A., Perret, L., Tillich, J.P.: A distinguisher for high-rate McEliece cryptosystems. IEEE Trans. Inform. Theory 59(10), 6830–6844 (2013)
[9] Faure, C., Minder, L.: Cryptanalysis of the McEliece cryptosystem over hyperelliptic codes. In: Proceedings 11th Int. Workshop on Algebraic and Combinatorial Coding Theory, ACCT 2008. pp. 99–107 (2008)
[10] Janwa, H., Moreno, O.: McEliece public crypto system using algebraic-geometric codes. Designs, Codes and Cryptography 8, 293–307 (1996)
[11] Kötter, R.: A unified description of an error locating procedure for linear codes. In: Proceedings of Algebraic and Combinatorial Coding Theory, pp. 113–117. Voneshta Voda (1992)
[12] Kötter, R.: On algebraic decoding of algebraic-geometric and cyclic codes. Ph.D. thesis, Linköping University of Technology, Linköping Studies in Science and Technology, Dissertation no. 419 (1996)
[13] Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R.: On the unique representation of very strong algebraic geometry codes. Designs, Codes and Cryptography pp. 215–230 (2012)
[14] Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R.: The non-gap sequence of a subcode of a generalized Reed-Solomon code. Designs, Codes and Cryptography 66(1-3) (2013)
[15] Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R., Ruano, D.: Computational aspects of retrieving a representation of an algebraic geometry code. J. Symbolic Computation 64, 67–87 (2014)
[16] Márquez-Corbella, I., Pellikaan, R.: Error-correcting pairs for a public-key cryptosystem. Preprint arXiv:1205.3647 (2012)
[17] McEliece, R.: A public-key cryptosystem based on algebraic coding theory. DSN Progress Report 42–44, 114–116 (1978)
[18] Minder, L.: Cryptography based on error correcting codes. Ph.D. thesis, 3846 EPFL (2007)
[19] Niederreiter, H.: Knapsack-type crypto systems and algebraic coding theory. Problems of Control and Information Theory 15(2), 159–166 (1986)
[20] Otmani, A.: Contribution to the cryptanalysis of code-based primitives. Université de Caen Basse Normandie, Caen (2011)
[21] Pellikaan, R.: On a decoding algorithm of codes on maximal curves. IEEE Trans. Inform. Theory 35, 1228–1232 (1989)
[22] Pellikaan, R.: On decoding by error location and dependent sets of error positions. Discrete Math. 106–107, 369–381 (1992)
[23] Pellikaan, R.: On the existence of error-correcting pairs. Statistical Planning and Inference 51, 229–242 (1996)
[24] Stichtenoth, H.: Algebraic function fields and codes. Springer, Berlin (1993)
[25] Tsfasman, M., Vlădut, S.: Algebraic-geometric codes. Kluwer Academic Publishers, Dordrecht (1991)
[26] V.M. Sidelnikov, V., Shestakov, S.: On the insecurity of cryptosystems based on generalized Reed-Solomon codes. Discrete Math. Appl. 2, 439–444 (1992)

Appendix A. The dimension of $A * B$

Let A and B be two linear codes over $\mathbb{F}_q$ with parameters $[n, s]$ and $[n, t]$, generated by the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_s\}$ and $\{\mathbf{b}_1, \ldots, \mathbf{b}_t\}$ of vectors in $\mathbb{F}_q^n$, respectively. Let M be an $st \times n$ matrix over $\mathbb{F}_q$ whose rows consist on the vectors $\mathbf{a}_i * \mathbf{b}_j = (a_{i,1}b_{j,1}, \ldots, a_{i,n}b_{j,n}) \in \mathbb{F}_q^n$ with $i \in \{1, \ldots, s\}$ and $j \in \{1, \ldots, t\}$ ordered lexicographically. Then the rows of M form a generating set of the code $A * B$.

Indeed M is a block-matrix consisting of s blocks $M_i = (\mathbf{a}_i * \mathbf{b}_j)_{1 \le j \le t}$ with $i \in \{1, \ldots, s\}$ of size $t \times n$. We define the support of a codeword $\mathbf{c} = (c_1, \ldots, c_n)$ by $\mathrm{supp}(\mathbf{c}) = \{i \mid c_i \neq 0\}$. Note that if $i \notin \mathrm{supp}(\mathbf{a}_j)$ then the i-th column of M_j consists of zeros.

In the following lines, assuming that $st < n$ we will prove that M has full rank with high probability. We proceed by a similar procedure as in [8, VI Theorem 2] where it is proved that the solution space of the linear system associated to an arbitrary random linear code is zero with high probability.

Let $E_1 = \mathrm{supp}(\mathbf{a}_1)$. Suppose $|E_1| \ge t$. Let F_1 be a subset of E_1 with cardinality t. To simplify notation and without loss of generality, we can always assume that F_1 corresponds to the first t elements in $\mathbf{a}_1$, by permuting the elements if necessary. Let $M^{(1)}$ be a square submatrix of M_1 formed by its first t columns, i.e.

$$M^{(1)} = (a_{1,j}b_{i,j})_{\substack{j \in F_1 \\ 1 \le i \le t}} \in \mathbb{F}_q^{t \times t}$$

Now we define by induction $E_i := \mathrm{supp}(\mathbf{a}_i) \setminus F_{i-1}$ and the subset F_i as the first t elements of the subset, assuming that $|E_i| \geq t$. The square matrix $M^{(i)} \in \mathbb{F}_q^{t\times t}$ is obtained from M_i by taking the F_i-indexed columns, for $i \in \{1, \ldots, s\}$. Then clearly the following Lemma holds.

Lemma Appendix A.1. *If $|E_i| \geq t$ for all $i \in \{1, \ldots, s\}$ then*

$$\mathrm{rank}(M) \geq \sum_{i=1}^{s} \mathrm{rank}(M^{(i)}).$$

Lemma Appendix A.2. *If $|E_i| \geq t$ for all $i = 1, \ldots, s$ then*

$$\Pr\left(\sum_{i=1}^{s} d(M^{(i)}) \geq u\right) \leq K^s q^{\frac{-u^2}{s}}$$

where $d(M^{(i)}) = t - \mathrm{rank}(M^{(i)})$ for $i = 1, \ldots, s$ and K is a constant depending only on q.

Proof. See [8, Lemma 5]. □

Lemma Appendix A.3. *Let $u_i = n - (i-1)t$ with $i = \{1, \ldots, s\}$, then*

$$\Pr\left(|E_i| < t, \mid |E_1| \geq t, \ldots, |E_{i-1}| \geq t\right) \leq e^{-2\frac{\left(\frac{q-1}{q}u_i - t + 1\right)^2}{u_i}}$$

Proof. See [7, Lemma 6] and [8]. □

Theorem Appendix A.4. *Assume that $st < n$. Then for any function $w(x)$ tending to infinity as x goes to infinity we have*

$$\Pr\left(D \geq w(t)\right) = o(1),$$

where $D = st - rank(M)$.

Proof. Note that if $|E_i| \geq t$ for $i \in \{1, \ldots, s\}$ then $D \leq \sum_{i=1}^{s} d(M^{(i)})$.

Let S_1 be the event $\sum_{i=1}^{s} d(M^{(i)}) \geq w(t)$ then using Lemma Appendix A.2 we have that $\Pr(S_1) = o(1)$. And let S_2 be the event of having at least one E_i with $i \in \{1, \ldots, s\}$ such that $|E_i| < t$. Then the probability of the complement of event S_2 is given by

$$\Pr\left(\overline{S_2}\right) = \Pr\left(\bigcap_{i=1}^{s} |E_i| \geq t\right) = \prod_{i=1}^{s} \Pr\left(|E_1| \geq t, \ldots, |E_i| \geq t\right) = 1 - o(1)$$

by Lemma Appendix A.3. Then we deduce that the sought probability is

$$\Pr\left(D \geq w(t)\right) \leq \Pr\left(S_1 \cup S_2\right) \leq \Pr\left(S_1\right) + \Pr\left(S_2\right) = o(1).$$

□