Error-correcting pairs for a public-key cryptosystem

Irene Márquez-Corbella and Ruud Pellikaan

Department of Algebra, Geometry and Topology, University of Valladolid
Facultad de Ciencias, 47005 Valladolid, Spain, E-mail: imarquez@agt.uva.es
Department of Mathematics and Computing Science, Eindhoven University of Techn.
P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: g.r.pellikaan@tue.nl

Abstract. Code-based cryptography is an interesting alternative to classic number-theory PKC since it is conjectured to be secure against quantum computer attacks. Many families of codes have been proposed for these cryptosystems, one of the main requirements is having high performance t-bounded decoding algorithms which in the case of having an error-correcting pair is achieved. In this article the class of codes with a t-ECP is proposed for the McEliece cryptosystem. The hardness of retrieving the t-ECP for a given code is considered. As a first step distinguishers of several subclasses are given.

Keywords: Code-based Cryptography, Error-Correcting Pairs.

1 Introduction

The notion of Public Key Cryptography (PKC) was first introduced in 1976 [10] by Diffie and Helman, though Merkle had previously developed some of the key concepts [32]. The main advantage with respect to symmetric-key cryptography is that it does not require an initial exchange of secrets between sender and receiver. In the survey paper [23] it is stated that

"At the heart of any public-key cryptosystem is a one-way function - a function $y = f(x)$ that is easy to evaluate but for which is computationally infeasible (one hopes) to find the inverse $x = f^{-1}(y)$".

The most famous trapdoor one-way functions are:

- **Integer factorization** where $x = (p, q)$ is a pair of prime numbers and $y = pq$ is its product. The best-known example of PKC is the Rivest-Shamir-Adleman (RSA) cryptosystem whose security is based on the hardness of distinguishing prime numbers from composite number, i.e. the intractability of inverting this one-way function.
- **Discrete logarithm** for which a group G (written multiplicatively) and an element $a \in G$ are required, then x is an integer and $y = a^x$. The security of the ElGamal cryptosystem or the Diffie-Hellman key exchange depends on the difficulty of finding discrete logarithms modulo a large prime.

– **Elliptic curve discrete logarithm** which it is actually a particular case of the previous function when G is taken as an elliptic curve group. Then $x = P$ is a point on the curve and $y = kP$ is another point on the curve obtained by the multiplication of P with a scalar k. Elliptic Curve Cryptography (ECC) proposed independently by Koblitz [22] and Miller [33] in 1985 is based on the difficulty of this function in the group of points on an elliptic curve over a finite field.

However with the discovery of Shor's algorithm [46] anyone with a quantum computer can break in polynomial time all cryptosystems whose security depends on the difficulty of the previous problem. Post-quantum cryptography gave birth to the next generation of cryptography algorithms, which are designed to run on conventional computers but no attacks by classical or quantum computers are known against them. See [5] for an overview of the state of the art in this area. Code-based cryptosystems such as McEliece [31] and Niederreiter [35] cryptosystems are interesting candidates for post-quantum cryptography. See the surveys [7, 13, 37, 44, 45].

The security of code-based cryptosystems is connected to the hardness of the general decoding problem which was shown by Berlekamp-McEliece-Van Tilborg [2, 4] to be NP-hard, even if preprocessing is allowed [8]. However it is not known whether this problem is almost always or in the average difficult. The problem of *minimum distance decoding* with input $(G, \mathbf{y})$ where G is a generator matrix of a code C over $\mathbb{F}_q$ of parameters $[n, k, d]$ addresses to determine a codeword $\mathbf{c} \in C$ of minimal distance to $\mathbf{y}$. The *bounded distance decoding problem* depends on a function $t(n, k, d)$. The input is again $(G, \mathbf{y})$ but the output is a codeword $\mathbf{c} \in C$ (if any) verifying that $d(\mathbf{y}, C) \leq t(n, k, d)$, where $d(\cdot, \cdot)$ denotes the hamming distance between two vectors on $\mathbb{F}_q^n$. Moreover *decoding up to half the minimum distance* is a bounded distance decoding problem such that $t(n, k, d) \leq \lfloor (d-1)/2 \rfloor$ for all n, k and d.

All known minimum distance decoding algorithm for general codes have exponential complexity in the length of the code. However there are several classes of codes such as the Reed-Solomon, BCH, Goppa or algebraic geometry codes which have polynomial decoding algorithms that correct up to a certain bound which is at most half the minimum distance.

The problems posed above have two parts [19]. Firstly the *preprocessing* part done at a laboratory or a factory where for an appropriate code C a decoder $\mathcal{A}_C$ is built which is allowed to be time consuming. Secondly the actual operating of the many copies of the decoder for consumers which should work very fast. So we can consider the problem of *minimum distance decoding with preprocessing*. From the error-correction point of view it seems pointless to decode a bad code, but for breaking the McEliece cryptosystem one must be able to decode efficiently all, or almost all, codes.

In 1978 [31] McEliece presents the first PKC based on the theory of error-correcting codes. Its main advantages are its fast encryption and decryption schemes. However the large key size of its public key makes it very difficult to use in many practical situations. In this cryptosystem the *public key space*

$\mathcal{K}$ is the collection of all generator matrices of a chosen class of codes that have an efficient decoding algorithm that corrects all patterns of t errors, the *plaintext space* is $\mathcal{P} = \mathbb{F}_q^k \times W_{n,q,t}$, where $W_{n,q,t}$ is the collection of all $\mathbf{e} \in \mathbb{F}_q^n$ of weight t, and the *ciphertext space* is $\mathcal{C} = \mathbb{F}_q^n$. The *sample space* is given by $\Omega = \mathcal{P} \times \mathcal{K}$. The *encryption map* $E_G : \mathcal{P} \to \mathcal{C}$ for a given key $G \in \mathcal{K}$ is defined by $E_G(\mathbf{m}, \mathbf{e}) = \mathbf{m}G + \mathbf{e}$. An *adversary* $\mathcal{A}$ is a map from $\mathcal{C} \times \mathcal{K}$ to $\mathcal{P}$. This adversary is successful for $(x, G) \in \Omega$ if $\mathcal{A}(E_G(x), G) = x$.

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 P. Then for the McEliece PKC the matrices G, S and P are kept secret while $G' = SGP$ is public. Furthermore the (trapdoor) one-way function of this cryptosystem is usually presented as follows:

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

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.

McEliece proposed to use the family of Goppa codes. The problem of bounded distance decoding for the class of codes that have the same parameters as the Goppa codes is difficult in the worst-case [15]. However, it is still an open problem whether decoding up to half the minimum distance is NP-hard which is the security basis of the McEliece cryptosystem. Algebraic geometry codes were also proposed for the McEliece PKC in [20, 34]. The security of this PKC is based on two assumptions [7, 21]:

A.1 In the average it is difficult to decode t errors for all codes that have the same parameters as the codes used as key,

A.2 It is difficult to distinguish arbitrary codes from those coming from $\mathcal{K}$.

Concerning the first assumption it might be that the class of codes is too small or too rigid. For instance Sidelnikov-Shestakov [50] gave an adversary that is always successful if one take for public key space the generator matrices of generalized Reed-Solomon (GRS) codes. Concerning the second assumption recent progress is made by Faugère et al. [14, 36] where they showed that one could distinguish between high rate Goppa, alternant and random codes.

In 1986 [35] Niederreiter presented a dual version of McEliece cryptosystem which is equivalent in terms of security [26]. Niederreiter's system differs from McEliece's system in the public-key structure (it use a parity check matrix instead of a generator matrix of the code), in the encryption mechanism (we compute the syndrome of a message by the public key) and in the decryption message. In its original paper Niederreiter proposed the class of GRS codes.

Let $H \in \mathbb{F}_q^{(n-k) \times n}$ be a parity check matrix of a code C in $\mathcal{C}$. H is masked by $H' = SHP$, where S is an invertible matrix over $\mathbb{F}_q$ of size $n - k$ and P is an $n \times n$ permutation or monomial matrix. The (trapdoor) one-way function in case of the Niederreiter PKC is presented by

$$x = \mathbf{e} \mapsto y = \mathbf{e}H'^T,$$

where $\mathbf{e} \in \mathbb{F}_q^n$ has weight t.

In a *syndrome based (SB) hash* function [1, 17, 16] an $n \times r$ parity check matrix H is chosen at random, then SB hash system is given by a procedure that encodes $\mathbf{s}$ bits of information into a word $\mathbf{e}$ of length n and weight t. The one-way function in this case (which has no trapdoor) is given by

$$x = \mathbf{e} \mapsto y = \mathbf{e}H^T.$$

It was shown in [12, 40, 41] that the known efficient bounded distance decoding algorithms of Reed-Solomon, BCH, Goppa and algebraic geometry codes can be described by a basic algorithm using an error correcting pair. That means that the proposed McEliece cryptosystem are not based on the inherent tractability of bounded distance decoding but on the one-way function

$$x = (A, B) \mapsto y = A * B,$$

where (A, B) is a t-error-correcting pair.

Consider $\mathcal{C}_t$, the class of linear codes over $\mathbb{F}_q$ that have a t-error correcting pair over an extension of $\mathbb{F}_q$. It was shown by Pellikaan [40] that codes of this family have an efficient decoding algorithm that corrects up to t errors. This makes them appropriate for code-based cryptography. Note that most families of codes used in such cryptosystems belong to $\mathcal{C}_t$ such as the generalized Reed-Solomon codes, the Goppa codes, the alternant codes and the algebraic-geometry codes.

For further details on the notion of error-correcting pair see Section 2 where we formally review this definition and we give a brief survey on the properties that are relevant to this work.

The aim of this paper is to study the subclass of $\mathcal{C}_t$ formed by those linear codes C whose error correcting pair is not easily reconstructed from C . Section 3 deals with the security status of this scheme, detailing the state-of-art and the existence of error-correcting pairs for families of codes most commonly used in code-based cryptography.

Finally, in Section 4, following the work of Faugère et al. [14], we present distinguishers for several families of codes. Recall that the hardness of the distinguishing problem was part of the basis of the security of code-based cryptosystems.

2 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 multiplication* is defined by coordinatewise multiplication, that is $\mathbf{a}*\mathbf{b} = (a_1b_1, \ldots, a_nb_n)$ while the *standard inner multiplication* is defined by $\mathbf{a}\cdot\mathbf{b} = \sum_{i=1}^n a_ib_i$. In general, for two subsets A and B of $\mathbb{F}_q^n$ the set $A*B$ is given by $\{\mathbf{a}*\mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B\}$. Furthermore $A \perp B$ if and only if $\mathbf{a} \cdot \mathbf{b} = 0$ for all $\mathbf{a} \in A$ and $\mathbf{b} \in B$.

Definition 1. *Let C be a linear code in $\mathbb{F}_q^n$. The pair (A, B) of linear subcodes of $\mathbb{F}_{q^m}^n$ is called a* t-error correcting pair *(ECP) 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 $d(A) + d(C) > n$.

The notion of an error-correcting pair for a linear code was introduced independently by Pellikaan in [38] and Kötter in [24]. In [38] 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 $\mathcal{O}(n^3)$. Furthermore the minimum distance of such linear code is at least $2t + 1$.

The existence of ECP for generalized Reed-Solomon and Algebraic codes was shown in [38] and for many cyclic codes Duursma and Kötter in [12] have found ECP which correct beyond the designed BCH capacity.

Note that if E.4 is replaced by the following statements

E.5 $d(A^{\perp}) > 1$ *i.e. A is a non-degenerated code,*
E.6 $d(A) + 2t > n$.

then $d(C) \geq 2t + 1$ and (A, B) is a t-ECP for C.

3 Error-correcting pairs for public key cryptosystems

Let $\mathcal{P}_t$ be the collection of pairs (A, B) such that A, B are linear codes over some extension of $\mathbb{F}_q$, A is non-degenerated and (A, B) is a t-error correcting pair for some linear code C in $\mathbb{F}_q^n$. We consider the following one way function

$$\begin{array}{rccc} \varphi: & \mathcal{P}_t & \longrightarrow & \mathbb{F}_q^n \\ & x = (A, B) & \longmapsto & y = A * B \end{array}$$

Let U and V be two generator matrices with rows denoted by $\mathbf{u}_i$ and $\mathbf{v}_i$, respectively, $U*V$ be the matrix form by the rows $\mathbf{u}_i*\mathbf{v}_j$ ordered lexicographically and $\mathrm{red}(U * V)$ be the matrix obtained from $U * V$ by deleting dependent rows. Then the implementation of φ may be given by

$$(U, V) \longmapsto y = \mathrm{red}(U * V)$$

Firstly we note that $\mathbf{u}P * \mathbf{v}P = (\mathbf{u} * \mathbf{v})P$ for every permutation or monomial matrix P. Thus, if (A, B) is a t-ECP for C, then (AP, BP) is a t-ECP for $P^{-1}C$. Furthermore, let S_1 and S_2 be invertible matrices of the correct sizes to be multiplied by the matrices U and V, respectively, then $U * V$ generates the same code as $(S_1U) * (S_2V)$ since $(S_1U) * \mathbf{v} = S_1(U * \mathbf{v})$ and $\mathbf{u} * (S_2V) = S_2(\mathbf{u} * V)$ for all vectors $\mathbf{u}$ and $\mathbf{v}$. Therefore the masking $H' = SHP$ 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).

Let C be the code with the elements of $A * B$ as parity checks. If the one-way function φ is indeed difficult to invert, then the code C with parity check matrix $H = \mathrm{red}(U * V)$ 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. In the following we consider seven collections of pairs.

Example 1. The class of GRS codes was proposed for code-based PKC by Niederreiter [35]. However this proposal is completely broken by the Sidelnikov-Shestakov attack given in [50].
Let $\mathbf{a}$ be an n-tuple of mutually distinct elements of $\mathbb{F}_q$ and 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

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

That is, if we define by induction $\mathbf{a}^1 = \mathbf{a}$ and $\mathbf{a}^{i+1} = \mathbf{a} * \mathbf{a}^i$, then $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is generated by the elements $\mathbf{b} * \mathbf{a}^i$ with $i = 0, \ldots, k-1$, i.e. 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 $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})^\perp = \mathrm{GRS}_{n-k}(\mathbf{a}, \mathbf{b}')$ for some $\mathbf{b}'$ 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 $A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{b}')$ and $B = \mathrm{GRS}_t(\mathbf{a}, \mathbf{1})$ is a t-ECP for C where $t = \left\lfloor \frac{n-k}{2} \right\rfloor$ and $\mathbf{b}' \in \mathbb{F}_q^n$ is a nonzero vector verifying that $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})^\perp = \mathrm{GRS}_{n-k}(\mathbf{a}, \mathbf{b}')$.
So GRS codes are the prime examples of codes that have a t-error-correcting pair. Moreover if C is an $[n, n-2t, 2t+1]$ code which has a t-error-correcting pair, then C is a generalized Reed-Solomon. This is trivial if $t = 1$, proved for $t = 2$ in [41, Theorem 6.5] and for arbitrary t in [30].

Example 2. Error-correcting pairs for cyclic codes were found by Duursma and Kötter [11, 12, 25]. Cyclic codes are not considered for applications in code-based PKC.

Example 3. The class of subcodes of GRS codes was proposed by Berger-Loidreau [3] 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 Wieschebrink [51, 52] and Márquez et al. [29].
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 1 which is also a t-ECP for C.

Example 4. Goppa codes were proposed for McEliece PKC by its author [31]. Sidelnikov-Shestakov made a claim [50] that their method for GRS codes could be extended to attack Goppa codes as well, but this was never substantiated by a paper in the public domain. In its original paper McEliece recommend the class of binary Goppa codes with parameters $[1024, 524, 101]$, but this proposal is no longer secure with nowadays computing power as shown in Peters et al.

[6, 42, 43] by improving decoding algorithms for general codes. The attack of Wieschebrink [52] is not yet efficient enough to be applicable to these codes.
A Goppa code associated to a Goppa polynomial of degree r can be viewed as an alternant code, that is a subfield subcode of GRS code of codimension r and therefore they have also a $\lfloor r/2 \rfloor$-error-correcting pair. In the binary case with an associated square free polynomial the Goppa code has an r-ECP.

Example 5. Algebraic geometry (AG) codes were introduced in 1977 by V.D. Goppa and were proposed by Janwa-Moreno [20] for the McEliece PKC. Recall that GRS codes can be seen as the class of AG codes on the projective line, i.e. the algebraic curve of genus zero. We refer the interested reader to [18, 48, 49].
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 m. 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 $m < n$, then the dimension of the code $C_L(\mathcal{X}, \mathcal{P}, E)$ is at least $m + 1 - g$ and its minimum distance is at least $n - m$. If $m > 2g - 2$, then its dimension is $m + 1 - g$. The dual code $C_L(\mathcal{X}, \mathcal{P}, E)^{\perp}$ is again AG. If $m > 2g - 2$, then the dimension of the code $C_L(\mathcal{X}, \mathcal{P}, E)^{\perp}$ is at least $n - m - 1 + g$ and its minimum distance is at least $d^* = m - 2g + 2$. If $m < n$, then its dimension is $n - m - 1 + g$.
If $A = C_L(\mathcal{X}, \mathcal{P}, E)$ and $B = C_L(\mathcal{X}, \mathcal{P}, F)$, then $\langle A * B \rangle \subseteq C_L(\mathcal{X}, \mathcal{P}, E + F)$. 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 [39, Theorem 1] and [40, Theorem 3.3]. If e is sufficiently large, then there exists a t-ECP over $\mathbb{F}_{q^e}$ with $t = \lfloor (d^* - 1)/2 \rfloor$ by [41, Proposition 4.2].
It was shown by Márquez et al. [28, 27] 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.
Geometric Goppa codes, which are subfield subcodes of Algebraic geometry codes [47] generalizing the classical Goppa codes that are subfield subcodes of GRS codes, were proposed for the McEliece PKC by Janwa-Moreno [20].

Example 6. If (A, B) is 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 [41, Corollary 3.4]. The dimension of $\langle A * B \rangle$ is at most $t(t + 1)$. So the dimension of C is at least $n - t(t + 1)$. In Appendix A it will be 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 code is MDS. So taking random codes A and B of length n and dimensions $t + 1$ and t, respectively, gives a very large class of code for the McEliece PKC. However with large field the key size becomes larger and recall that the main obstacle for coded-based crypto systems was the key size.

Example 7. If (A, B) is a pair of codes that satisfy the conditions (E.1), (E.2), (E.3), (E.5) and (E.6), then the minimum distance of C is at least $2t + 1$ and (A, B) is a t-error-correcting pair for C by [41, Corollary 3.4].

4 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 by $\langle C * C\rangle$ or $C^{(2)}$ 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 [9, §4 Definition 6] and [29, 52]. Now, following the same scheme as in [28], we consider the linear map

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

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 [14, §3] and [36, 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

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

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 1.

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

Proof. Let M be the $\binom{k+1}{2} \times n$ matrix with entries $(g_{ij}g_{i'j})_{\substack{1\le i\le i'\le k\\ 1\le j\le n}}$. 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

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

Let M_1 be the $k \times \binom{n-k}{2}$ matrix with entries $(p_{ij}p_{ij'})_{\substack{1\le i\le k\\ k<j<j'\le n}}$, 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 $(h_{ij}h_{i'j})_{\substack{1\le i\le i'\le n-k\\ 1\le j\le n}}$. 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 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:

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

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

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

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 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. [14] 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. □

Example 8. Let C be a GRS code with parameters $[n,k]$, take for instance $C = \text{GRS}_k(\mathbf{a},\mathbf{b})$ 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 $\text{GRS}_{2k-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$ 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 9. Let C be a k-dimensional subcode of the code $\text{GRS}_l(\mathbf{a},\mathbf{b})$. Then $C^{(2)}$ is a subcode of the code $\text{GRS}_{2l-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$, 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 [29] that $C^{(2)}$ is equal to $\text{GRS}_{2l-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$ 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 $\text{GRS}_l(\mathbf{a},\mathbf{b})^\perp = \text{GRS}_{n-l}(\mathbf{a},\mathbf{b}')$. That is, $D^{(2)}$ contains the square of $\text{GRS}_{n-l}(\mathbf{a},\mathbf{b}')$ which is equal to $\text{GRS}_{2n-2l-1}(\mathbf{a},\mathbf{b}'*\mathbf{b}')$ 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 $\text{GRS}_{2n-2l-1}(\mathbf{a},\mathbf{b}'*\mathbf{b}') \subseteq D^{(2)}$ 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 10. The problem of distinguishing Goppa, alternant and random codes from each other was studied by Faugère et al. in [14]. 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 11. 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 m. Then $C^{(2)} \subseteq \mathcal{C}_L(\mathcal{X},\mathcal{P},2E)$. Assume moreover that $2g-2 < m < n/2$. Then C has dimension $k = m+1-g$

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

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

where $\{f_0, \ldots, f_{m-g}\}$ 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 $2g + 2 \leq m < \frac{1}{2}n$, 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

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

Therefore $\mathcal{Y}$ is given as the intersection of at least $\binom{k}{2} - m$ quadrics in $\mathbb{P}^{m-g}$. For more details we refer the reader to [28].

Example 12. 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 6. Then $D = C^{\perp} = \langle A * B \rangle$. So $D^{(2)} = \langle A^{(2)} * B^{(2)} \rangle$. Hence

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

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

References

1. Augot, D., Finiasz, M., Sendrier, N.: A Family of Fast Syndrome Based Cryptographic Hash Function. In: Dawson, E., Vaudenay, S. (eds.) MyCrypt 2005, Lecture Notes in Computer Science. vol. 3715, pp. 64–83. Springer, Berlin (2005)
2. Barg, A.: Complexity issues in coding theory. In: Pless, V., Huffman, W. (eds.) Handbook of coding theory, vol. 1, pp. 649–754. North-Holland, Amsterdam (1998)
3. Berger, T., Loidreau, P.: How to mask the structure of codes for a cryptographic use. Designs, Codes and Cryptography 35, 63–79 (2005)
4. Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems. IEEE Transactions on Information 24, 384–386 (1978)
5. Bernstein, D.: Introduction to post-quantum cryptography. In: D.J. Bernstein, J.B., Dahmen, E. (eds.) Post-quantum cryptography, pp. 1–14. Springer-Verlag, Berlin (2009)
6. Bernstein, D., Lange, T., Peters, C.: Attacking and defending the McEliece cryptosystem. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 5299, pp. 31–46. Springer, Berlin (2008)

7. Biswas, B., Sendrier, N.: McEliece cryptosystem implementation : Theory and practice. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 5299, pp. 47–62. Springer, Berlin (2008)
8. Bruck, J., Naor, N.: The hardness of decoding linear codes with preprocessing. IEEE Transactions on Information 36, 381–385 (1990)
9. 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)
10. Diffie, W., Hellman, M.E.: New directions in cryptography. In: Secure communications and asymmetric cryptosystems, AAAS Sel. Sympos. Ser., vol. 69, pp. 143–180. Westview, Boulder, CO (1982)
11. Duursma, I.: Decoding codes from curves and cyclic codes. Ph.D. thesis, Eindhoven University of Technology (1993)
12. Duursma, I., Kötter, R.: Error-locating pairs for cyclic codes. IEEE Trans. Inform. Theory 40, 1108–1121 (1994)
13. Engelbert, D., Overbeck, R., Schmidt, A.: A summary of McEliece-type cryptosystems and their security. Journal of Mathematical Cryptology 1(2), 151–199 (2007)
14. 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)
15. Finiasz, M.: Nouvelles constructions utilisant des codes correcteurs d'erreurs en cryptographie a clef publique. Ph.D. thesis, INRIA - Ecole Polytechnique (2004)
16. Finiasz, M.: Syndrome based collision resistant hashing. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 5299, pp. 137–147. Springer, Berlin (2008)
17. Finiasz, M., Gaborit, P., Sendrier, N.: Improved fast syndrome based cryptographic hash function. In: Proceedings of ECRYPT Hash Workshop 2007. Barcelona (2007)
18. Goppa, V.: Codes associated with divisors. Probl. Inform. Transmission 13, 22–26 (1977)
19. Høholdt, T., Pellikaan, R.: On decoding algebraic-geometric codes. IEEE Transactions on Information 41, 1589–1614 (1995)
20. Janwa, H., Moreno, O.: McEliece public crypto system using algebraic-geometric codes. Designs, Codes and Cryptography 8, 293–307 (1996)
21. Kobara, K., Imai, H.: Semantically secure McEliece public-key cryptosystems - conversions for McEliece PKC. In: PKC 2001, Lecture Notes in Computer Science. vol. 1992, pp. 19–35. Springer, Berlin (2001)
22. Koblitz, N.: Elliptic curve cryptosystems. Mathematics of Computation 48(177), 203–209 (1987)
23. Koblitz, N., Menezes, A.: The brave new world of bodacious assumptions in cryptography. Notices Amer. Math.Soc. 57(3), 357–365 (2010)
24. 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)
25. 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)
26. Li, Y., Deng, R., Wang, X.: The equivalence of McEliece's and Niederreiter's public-key cryptosystems. IEEE Transactions on Information 40, 271–273 (1994)

27. Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R.: Cryptanalysis of public-key cryptosystems based on algebraic geometry codes. Submitted to Designs, Codes and Cryptographie (November 2011)
28. Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R.: Evaluation of public-key cryptosystems based on algebraic geometry codes. In: Proceedings of the Third International Castle Meeting on Coding Theory and Applications. pp. 199–204. Cardona Castle, Barcelona (September 11-15, 2011)
29. Márquez-Corbella, I., Martínez Moro, E., Pellikaan, R.: The non-gap sequence of a subcode of a generalized Reed-Solomon code,. In: Proceedings of the Seventh International Workshop on Coding and Cryptography, WCC 2011. pp. 183–193. Paris (2011)
30. Márquez-Corbella, I., Pellikaan, R.: A characterization of MDS codes that have an error-correcting pair. Preprint
31. McEliece, R.: A public-key cryptosystem based on algebraic coding theory. DSN Progress Report 42–44, 114–116 (1978)
32. Merkle, R.: Secure communications over insecure channels. In: Secure communications and asymmetric cryptosystems, AAAS Sel. Sympos. Ser., vol. 69, pp. 181–196. Westview, Boulder, CO (1982)
33. Miller, V.: Use of Elliptic Curves in Cryptography. In: Williams, H. (ed.) Advances in Cryptology - CRYPTO 85 Proceedings, Lecture Notes in Computer Science, vol. 218, pp. 417–426. Springer Berlin / Heidelberg (1986)
34. Niebuhr, R.: Public key cryptography based on coding theory. Master's thesis, Tech. Univ. Darmstadt (2006)
35. Niederreiter, H.: Knapsack-type crypto systems and algebraic coding theory. Problems of Control and Information Theory 15(2), 159–166 (1986)
36. Otmani, A.: Contribution to the cryptanalysis of code-based primitives. Université de Caen Basse Normandie, Caen (2011)
37. Overbeck, R.: Public key cryptography based on coding theory. Master's thesis, Tech.Univ. Darmstadt (2007)
38. Pellikaan, R.: On decoding linear codes by error correcting pairs (1988), preprint Eindhoven University of Technology
39. Pellikaan, R.: On a decoding algorithm of codes on maximal curves. IEEE Trans. Inform. Theory 35, 1228–1232 (1989)
40. Pellikaan, R.: On decoding by error location and dependent sets of error positions. Discrete Math. 106–107, 369–381 (1992)
41. Pellikaan, R.: On the existence of error-correcting pairs. Statistical Planning and Inference 51, 229–242 (1996)
42. Peters, C.: Information-set decoding for linear codes over Fq. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 6061, pp. 81–94. Springer, Berlin (2010)
43. Peters, C.: Curves, codes and cryptography. Ph.D. thesis, Technical University Eindhoven (2011)
44. Sendrier, N.: McEliece public key cryptosystem. In: van Tilborg, H. (ed.) Encyclopedia of cryptography and security, pp. 375–378. Springer, Berlin (2005)
45. Sendrier, N.: Niederreiter encryption scheme. In: van Tilborg, H. (ed.) Encyclopedia of cryptography and security, pp. 413–414. Springer, Berlin (2005)
46. Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J. Comput. 26, 1484–1509 (October 1997)
47. Skorobogatov, A.: The parameters of subcodes of algebraic-geometric codes over prime subfields. Discrete Appl. Math. 33, 205–214 (1991)

48. Stichtenoth, H.: Algebraic function fields and codes. Springer, Berlin (1993)
49. Tsfasman, M., Vlădut, S.: Algebraic-geometric codes. Kluwer Academic Publishers, Dordrecht (1991)
50. V.M. Sidelnikov, V., Shestakov, S.: On the insecurity of cryptosystems based on generalized Reed-Solomon codes. Discrete Math. Appl. 2, 439–444 (1992)
51. Wieschebrink, C.: An attack on the modified Niederreiter encryption scheme. In: PKC 2006, Lecture Notes in Computer Science. vol. 3958, pp. 14–26. Springer, Berlin (2006)
52. Wieschebrink, C.: Cryptanalysis of the Niederreiter public key scheme based on GRS subcodes. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 6061, pp. 61–72. Springer, Berlin (2010)

A The dimension of $\langle A * B \rangle$

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 $M_i = (\mathbf{a}_i * \mathbf{b}_j)_{1 \leq j \leq t}$ 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 on 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 Appendix B of [14] 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 $|E_1| \geq t$. 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.

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

Now we define by induction $E_i := \mathrm{supp}(\mathbf{a}_i) \backslash F_{i-1}$ 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 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 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 [14, Lemma 5]. □

Lemma 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 [14, Lemma 6]. □

Theorem 1. *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 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 3.

Then we deduce that the sought probability is

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

□