Algebraic Cryptanalysis of Compact McEliece's Variants – Toward a Complexity Analysis

Jean-Charles Faugère[1], Ayoub Otmani[2,3], Ludovic Perret[1], and Jean-Pierre Tillich[2]

[1] SALSA Project - INRIA (Centre Paris-Rocquencourt)
UPMC, Univ Paris 06 - CNRS, UMR 7606, LIP6
104, avenue du Président Kennedy 75016 Paris, France
jean-charles.faugere@inria.fr, ludovic.perret@lip6.fr
[2] SECRET Project - INRIA Rocquencourt
Domaine de Voluceau, B.P. 105 78153 Le Chesnay Cedex - France
ayoub.otmani@inria.fr, jean-pierre.tillich@inria.fr
[3] GREYC - Université de Caen - Ensicaen
Boulevard Maréchal Juin, 14050 Caen Cedex, France.

Abstract. A new algebraic approach to investigate the security of the McEliece cryptosystem has been proposed by Faugère-Otmani-Perret-Tillich in Eurocrypt 2010. This paper is an extension of this work. The McEliece's scheme relies on the use of error-correcting codes. It has been proved that the private key of the cryptosystem satisfies a system of bi-homogeneous polynomial equations. This property is due to the particular class of codes considered which are alternant codes. These highly structured algebraic equations allowed to mount an efficient key-recovery attack against two recent variants of the McEliece cryptosystems that aim at reducing public key sizes by using quasi-cyclic or quasi-dyadic structures. Thanks to a very recent development due to Faugère-Safey el Din-Spaenlehauer on the solving of bihomogeneous bilinear systems, we can estimate the complexity of the FOPT algebraic attack. This is a first step toward providing a concrete criterion for evaluating the security of future compact McEliece variants.

Keywords : public-key cryptography, McEliece cryptosystem, algebraic cryptanalysis, F_5, bi-linear systems,.

1 Introduction

One of the main goals of the public-key cryptography is the design of secure encryption schemes by exhibiting one-way trapdoor functions. This requires the identification of supposedly hard computational problems. Although many hard problems exist and are proposed as a foundation for public-key primitives, those effectively used are essentially classical problems coming from number theory: integer factorization (e.g. in RSA) and discrete logarithm (e.g. in Diffie-Hellman key-exchange). However, the lack of diversity in public key cryptography is a major concern in the field of information security. This situation would worsen if ever quantum computers appear because schemes that are based on these classical number theory problems would become totally insecure.

Consequently, the task of identifying alternative hard problems that are not based on number theory ones constitutes a major issue in the modern public-key cryptography. Among those problems, the intractability of decoding a random linear code [7] seems to offer the most promising solution thanks to McEliece who first proposed in [24] a public-key cryptosystem based on irreducible binary Goppa codes. The class of Goppa codes represents one of the most important example of linear codes having an efficient decoding algorithm [8, 27]. The resulting cryptosystem has then very fast encryption and decryption functions [10]. A binary Goppa code is defined by a polynomial $g(z)$ of degree $r \geq 1$ with coefficients in some extension $\mathbb{F}_{2^m}$ of degree $m > 1$ over $\mathbb{F}_2$, and a n-tuple $\mathscr{L} = (x_1, \ldots, x_n)$ of distinct elements in $\mathbb{F}_{2^m}$ with $n \leq 2^m$. The trapdoor of the McEliece public-key scheme consists of the randomly picked $g(z)$ with $\mathscr{L}$ which together provide all the information to decode efficiently. The public key is a randomly picked generator matrix of the chosen Goppa code. A ciphertext is obtained by multiplying a plaintext with the public generator matrix and adding a random error vector of prescribed Hamming weight. The receiver decrypts the message thanks to the decoding algorithm that can be derived from the secrets.

After more than thirty years now, the McEliece cryptosystem still belongs to the very few public key cryptosystems which remain unbroken. Its security relies upon two assumptions: the *intractability of decoding random linear codes* [7], and the *difficulty of recovering the private key* or an equivalent one. The problem of decoding an unstructured code is a long-standing problem whose most effective algorithms [20, 21, 29, 12, 9] have an exponential time complexity. On the other hand no significant breakthrough has been observed during the past years regarding the problem of recovering the private key. Indeed, although some weak keys have been identified in [22], the only known key-recovery attack is the exhaustive search of the secret polynomial $\Gamma(z)$ of the Goppa code, and applying the *Support Splitting Algorithm* (SSA) [28] to check whether the Goppa code candidate is *permutation-equivalent* to the code defined by the public generator matrix.

Despite its impressive resistance against a variety of attacks and its fast encryption and decryption, McEliece cryptosystem has not stood up to RSA for practical applications. This is most likely due to the large size of the public key which is between several hundred thousand and several million bits. To overcome this limitation, a trend had been initiated in order to decrease the key size by focusing on very structured codes. For instance, quasi-cyclic code like in [19], or quasi-cyclic codes defined by sparse matrices (also called LDPC codes) [1]. Both schemes were broken in [26]. It should be noted that the attacks have no impact on the security of the McEliece cryptosystem since both proposals did not use the binary Goppa codes of the McEliece cryptosystem. These works were then followed by two independent proposals [6, 25] that are based on the same kind of idea of using quasi-cyclic [6] or quasi-dyadic structure [25]. These two approaches were also broken in [18] where for the first time an algebraic attack is introduced against the McEliece cryptosystem.

Algebraic cryptanalysis is a general framework that permits to assess the security of theoretically all cryptographic schemes. So far, such type of attacks has been applied successfully against several multivariate schemes and stream ciphers. The basic principle of this cryptanalysis is to associate to a cryptographic primitive a set of algebraic equations. The system of equations is constructed in such a way to have a correspondence between the solutions of this system, and a secret information of the cryptographic primitive (for instance the secret key of an encryption scheme). In the case of the McEliece cryptosystem, the algebraic system that has to be solved has the following very specific structure:

$$\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y}) = \left\{ g_{i,0}Y_0X_0^j + \cdots + g_{i,n-1}Y_{n-1}X_{n-1}^j = 0 \;\middle|\; i \in \{0,\ldots,k-1\}, j \in \{0,\ldots,r-1\} \right\} \tag{1}$$

where the unknowns are the X_i's and the Y_i's and the $g_{i,j}$'s are known coefficients with $0 \leq i \leq k-1, 0 \leq j \leq n-1$ that belong to a certain field $\mathbb{F}_q$ with $q = 2^s$. We look for solutions of this system in a certain extension field $\mathbb{F}_{q^m}$. Here k is an integer which is at least equal to $n - rm$. By denoting $\mathbf{X} \overset{\text{def}}{=} (X_0, \ldots, X_{n-1})$ and $\mathbf{Y} \overset{\text{def}}{=} (Y_0, \ldots, Y_{n-1})$ we will refer to such an algebraic system by $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$. This algebraic approach as long as the codes that are considered are alternant codes. It is important to note that a Goppa code can also be seen as a particular alternant code. However, it is not clear whether an algebraic attack can be mounted efficiently against the original McEliece cryptosystem because the total number of equations is rk, the number of unknowns $2n$ and the maximum degree $r-1$ of the equations can be extremely high (e.g. $n = 1024$ and $r - 1 = 49$).

But in the case of the tweaked McEliece schemes [6, 25], it turns out that is possible to make use of this structure in order to reduce considerably the number of unknowns in the algebraic system. This is because of the type of codes that are considered: quasi-cyclic alternant codes in [6] and quasi-dyadic Goppa codes in [25]. In particular, it induces an imbalance between the $\mathbf{X}$ and $\mathbf{Y}$ variables. Moreover, it was possible to solve efficiently the algebraic system thanks to a dedicated Gröbner bases techniques. Finally, it was also observed experimentally in [18] but not formally proved that the complexity of the attack is mainly determined by the number of remaining variables in the block $\mathbf{Y}$.

The motivation of this paper is to revisit the FOPT algebraic attack [18] in view of the recent results on bilinear systems [16]. This permits to make more precise the dependency between the security of a McEliece (and its variants) and the properties of the algebraic system (1). This is a first step toward providing a concrete criterion for evaluating the security of future compact McEliece variants.

Organisation of the paper. After this introduction, the paper is organized as follows. We briefly recall the McEliece cryptosystem in the next section. In Section 3, we recall how we can derive the algebraic system (1). We emphasize that these parts are similar to the ones in [18]. Section 4 is the core of the paper. We explain how we can extract a suitable (i.e. affine bi-linear) system from $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$. We then recall new results on the complexity of solving generic affine bi-linear systems, which permit to obtain a rough estimate of the complexity of the FOPT attack. Finally, in Section 6, we compare our theoretical bound with the practical results obtained in [18].

2 McEliece Public-Key Cryptosystem

We recall here how the McEliece public-key cryptosystem is defined.

Secret key: the triplet (S, G_s, P) of matrices defined over a finite field $\mathbb{F}_q$ over q elements, with q being a power of two, that is $q = 2^s$. G_s is a full rank matrix of size $k \times n$, with $k < n$, S is of size $k \times k$ and is invertible, and P is permutation matrix of size $n \times n$. Moreover G_s defines a code (which is the set of all possible uG_s with u ranging over $\mathbb{F}_q^k$) which has a decoding algorithm which can correct in polynomial time a set of errors of weight at most t. This means that it can recover in polynomial time u from the knowledge of $uG_s + e$ for all possible $e \in \mathbb{F}_q^n$ of Hamming weight at most t.

Public key: the matrix product $G = SG_sP$.

Encryption: A plaintext $u \in \mathbb{F}_q^k$ is encrypted by choosing a random vector e in $\mathbb{F}_q^n$ of weight at most t. The corresponding ciphertext is $c = uG + e$.

Decryption: $c' = cP^{-1}$ is computed from the ciphertext c. Notice that $c' = (uSG_sP + e)P^{-1} = uSG_s + eP^{-1}$ and that eP^{-1} is of Hamming weight at most t. Therefore the aforementioned decoding algorithm can recover in polynomial time uS. This vector is multiplied by S^{-1} to obtain the plaintext u.
This describes the general scheme suggested by McEliece. From now on, we say that G is the *public generator matrix* and the vector space $\mathscr{C}$ spanned by its rows is the *public code i.e.* $\mathscr{C} \overset{\text{def}}{=} \{uG \mid u \in \mathbb{F}_q^k\}$. What is generally referred to as the McEliece cryptosystem is this scheme together with a particular choice of the code, which consists in taking a binary Goppa code. This class of codes belongs to a more general class of codes, namely the alternant code family ([23, Chap. 12, p. 365]). The main feature of this last class of codes is the fact that they can be decoded in polynomial time.

3 McEliece's Algebraic System

In this part, we explain more precisely how we construct the algebraic system described in (1). As explained in the previous section, the McEliece cryptosystem relies on Goppa codes which belong to the class of *alternant codes* and inherit from this an efficient decoding algorithm. It is convenient to describe such codes through a *parity-check matrix*. This is an $r \times n$ matrix H defined – over an extension $\mathbb{F}_{q^m}$ of the field where the code is constructed – as follows:

$$\{uG_s \mid u \in \mathbb{F}_q^k\} = \{c \in \mathbb{F}_q^n \mid Hc^T = 0\}. \tag{2}$$

r satisfies in this case the condition $r \geq \frac{n-k}{m}$. For alternant codes, there exists a parity-check matrix with a very special form related to Vandermonde matrices. More precisely there exist two vectors $x = (x_0, \ldots, x_{n-1})$ and $y = (y_0, \ldots, y_{n-1})$ in $\mathbb{F}_{q^m}^n$ such that $V_r(x,y)$ is a parity-check matrix, with

$$V_r(x,y) \overset{\text{def}}{=} \begin{pmatrix} y_0 & \cdots & y_{n-1} \\ y_0x_0 & \cdots & y_{n-1}x_{n-1} \\ \vdots & & \vdots \\ y_0x_0^{r-1} & \cdots & y_{n-1}x_{n-1}^{r-1} \end{pmatrix}. \tag{3}$$

We use the following notation in what follows.

Definition 1. *The alternant code $\mathscr{A}_r(x,y)$ of order r over $\mathbb{F}_q$ associated to $x=(x_0,\dots,x_{n-1})$ where the x_i's are different elements of $\mathbb{F}_{q^m}$ and $y=(y_0,\dots,y_{n-1})$ where the y_i's are nonzero elements of $\mathbb{F}_{q^m}$ is defined by $\mathscr{A}_r(x,y)=\{c\in\mathbb{F}_q^n \mid V_r(x,y)c^T=0\}$.*

It should be noted that the public code in the McEliece scheme is also an alternant code. We denote here by the public code, the set of vectors of the form

$$\{uG \mid u\in\mathbb{F}_q^k\}=\{cSG_sP \mid c\in\mathbb{F}_q^k\}.$$

This is simple consequence of the fact that the set $\{uSG_sP \mid u\in\mathbb{F}_q^k\}$ is obtained from the secret code $\{uG_s \mid u\in\mathbb{F}_q^k\}$ by permuting coordinates in it with the help of P, since multiplying by an invertible matrix S of size $k\times k$ leaves the code globally invariant. The key feature of an alternant code is the following fact.

Fact 1. *There exists a polynomial time algorithm decoding an alternant code once a parity-check matrix H of the form $H=V_r(x,y)$ is given.*

In other words, it is possible to break the McEliece scheme once we can find x^* and y^* in $\mathbb{F}_{q^m}^n$ such that

$$\{xG \mid x\in\mathbb{F}_q^n\}=\{y\in\mathbb{F}_q^n \mid V_r(x^*,y^*)y^T=0\}. \tag{4}$$

From the knowledge of this matrix $V_r(x^*,y^*)$, it is possible to decode the public code, that is to say to recover u from $uG+e$. Finding such a matrix clearly amounts to find a matrix $V_r(x^*,y^*)$ such that $V_r(x^*,y^*)G^T=0$. Let $X_0,\dots,X_{n-1}$ and $Y_0,\dots,Y_{n-1}$ be $2n$ variables corresponding to the x_i^*s and y_i^* respectively. We see that finding such values is equivalent to solve the following system:

$$\left\{ g_{i,0}Y_0X_0^j+\cdots+g_{i,n-1}Y_{n-1}X_{n-1}^j=0 \;\middle|\; i\in\{0,\dots,k-1\}, j\in\{0,\dots,r-1\} \right\} \tag{5}$$

where the $g_{i,j}$'s are the entries of the known matrix G with $0\le i\le k-1$ and $0\le j\le r-1$.

The cryptosystems proposed in [6, 25] follow the McEliece scheme [24] with the additional goal to design a public-key cryptosystem with very small key sizes. They both require to identify alternant codes having a property that allows matrices to be represented by very few rows. In the case of [6] circulant matrices are chosen whereas the scheme [25] focuses on dyadic matrices. These two families have in common the fact the matrices are completely described from the first row. The public generator matrix G in these schemes is a block matrix where each block is circulant in [6] and dyadic in [25]. The algebraic approach previously described leaded to a key-recovery in nearly all the parameters proposed in both schemes [18]. The crucial point that makes the attack possible is due to the very particular structure of the matrices and their block form describing the public alternant codes. This permits to drastically reduce the number of variables in $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$.

4 On Solving $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$

Thanks to a very recent development [16] on the solving of bi-linear systems, we can revisit the strategy used in [18] to solve $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$. As we will see, this permits to evaluate the complexity of computing a Gröbner bases of $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$ for compact variants of McEliece such as [6, 25]. Before that, we recall basic facts about the complexity of computing Gröbner bases [11, 13–15].

4.1 General Complexity of Gröbner Bases

The complexity of computing such bases depends on the so-called *degree of regularity*, which can be roughly viewed as the maximal degree of the polynomials appearing during the computation. This degree of regularity, denoted D_{reg} in what follows, is the key parameter. Indeed, the cost of computing a Gröbner basis is polynomial in the degree of regularity D_{reg}. Precisely, the complexity is:

$$\mathscr{O}\left(\binom{N+D_{\text{reg}}}{D_{\text{reg}}}^{\omega}\right), \tag{6}$$

which basically correspond to the complexity of reducing a matrix of size $\binom{N+D_{\text{reg}}}{D_{\text{reg}}}$ ($2 < \omega \leq 3$ is the "linear algebra constant", and N the number of variables of the system). The behavior of the degree of regularity D_{reg} is well understood for random (i.e. regular and semi-regular) systems [2, 4, 3, 5].

Proposition 1. *The degree of regularity of a square regular quadratic system in* $\mathbf{X}$ *is bounded by:*

$$1 + n_X, \tag{7}$$

where n_X is the number of variables in the set of variables $\mathbf{X}$*. Consequently, the maximal degree occurring in the computation of a DRL Gröbner basis (Degree Reverse Lexicographical order see [13]) is bounded by the same bound* (7)*.*

On the contrary, as soon as the system has some kind of structure as for $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$, this degree is much more difficult to predict in general. Typically, It is readily seen that $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$ has a very specific structure, it is bi-homogeneous (i.e. product of two homogeneous polynomials with distinct variables).

4.2 Extracting a Bi-Affine System from $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$

As explained, $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$ is highly structured. It is very sparse as the only monomials occurring in the system are of the form $Y_i X_i^j$, with $0 \leq i \leq k-1$ and $0 \leq j \leq r-1$. It can also be noticed that each block of k equations is *bi-homogeneous*, i.e. homogeneous if the variables of $\mathbf{X}$ (resp. $\mathbf{Y}$) are considered alone. More precisely, we shall say that $f \in \mathbb{F}_{q^m}[\mathbf{X},\mathbf{Y}]$ is *bi-homogeneous* of *bi-degree* (d_1,d_2) if:

$$\forall \alpha, \mu \in \mathbb{F}_{q^m},\ f(\alpha\mathbf{X}, \mu\mathbf{Y}) = \alpha^{d_1}\mu^{d_2} f(\mathbf{X}, \mu\mathbf{Y}).$$

Note that the equations occurring in $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$ are of bi-degree $(j,1)$, with $j, 0 \leq j \leq r-1$.

We briefly recall now the strategy followed in [18] to solve $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$. The first fundamental remark is that there are k linear equations in the n variables of the block $\mathbf{Y}$ in $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$. This implies that all the variables of the block $\mathbf{Y}$ can be expressed in terms of $n_{Y'} \geq n-k$ variables. From now on, we will always assume that the variables of the block $\mathbf{Y}'$ only refer to these $n_{Y'}$ *free* variables. The first step is then to rewrite the system (1) only in function of the variables of $\mathbf{X}$ and $\mathbf{Y}'$, i.e., the variables of $\mathbf{Y} \setminus \mathbf{Y}'$ are substituted by linear combinations involving only variables of $\mathbf{Y}'$.

In the particular cases of [6, 25], the quasi-cyclic and dyadic structures provide additional linear equations in the variables of $\mathbf{X}$ and $\mathbf{Y}'$ which can be also used to rewrite/clean the system. In the sequel, we denote by $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ the system obtained from $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$ by removing all the linear equations in $\mathbf{X}$ and $\mathbf{Y}$.

This system $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ being naturally overdetermined, we can "safely" remove some equations. In [6, 25], the system $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ is always defined over a field of characteristic two. It makes sense then to consider the set of equations of $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ whose degree in the variables of $\mathbf{X}'$ is a power of 2, i.e. equations of bi-degree $(2^j,1)$. We obtain in this way a sub-system of $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$, denoted $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$, having $n_{X'}$ and $n_{Y'}$ variables and at most $k \cdot \log_2(r)$ equations. This system is a "quasi" bi-linear system over $\mathbb{F}_2^m$ as $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ viewed over $\mathbb{F}_2$ is bi-linear. Note that some constant terms can occur in $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$, so the system is more precisely *affine* bi-linear.

Proposition 2. *Let* $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}') \subset \mathbb{F}_{q^m}[\mathbf{X}',\mathbf{Y}']$ *be the system from* $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ *by considering only the equations of bi-degree* $(2^j,1)$*. This system has $n_{X'} + n_{Y'}$ variables, at most $k \cdot \log_2(r)$ equations and is affine bi-linear.*

4.3 On the Complexity of Solving Affine Bi-Linear Systems

Whilst the complexity of solving general bi-homogenous system is not known, the situation is different for bi-affine (resp. bi-linear) systems. In particular, the theoretical complexity is well mastered, and there is a now a dedicated algorithm for such systems [16]. As already explained, our equations are "quasi" bi-linear as we are working with equations of bi-degree $(1,2^j)$ over a field of characteristic 2. The results presented in [16] can be then extended with a slight adaptation to the context.

A first important result of [16] is that F$_5$ [15] algorithm is already optimal for "generic" (random) affine bi-linear systems, i.e. all reductions to zero are removed by the F$_5$ criterion. Another fundamental result is that the degree of regularity of a square generic affine bi-linear system is much smaller than the degree of regularity of a generic system. It has been proved [16] that:

Proposition 3. *The degree of regularity of a square generic affine bi-linear system in* $\mathbf{X}$*' and* $\mathbf{Y}$*' is bounded by:*

$$1+\min(n_{X'},n_{Y'}), \tag{8}$$

where $n_{X'}$ *and* $n_{Y'}$ *are the number of variables in the blocks* $\mathbf{X}'$ *and* $\mathbf{Y}'$ *respectively. Consequently, the maximal degree occurring in the computation of a DRL Gröbner basis is also bounded by* (8)*.*

Remark 1. This bound is sharp for a generic square affine bi-linear system and is much better than the usual Macaulay's bound (7) for a similar quadratic system (that is to say a system of $n_{X'}+n_{Y'}$ quadratic equations in $n_{X'}+n_{Y'}$ variables):

$$1+\min(n_{X'},n_{Y'}) \ll 1+n_{X'}+n_{Y'}$$

Since $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ is a bilinear system it is reasonable to derive a bound for this system from the previous result:

Proposition 4. *Let* $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ *be as defined below. The maximum degree reached when computing a Gröbner basis of* $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ *is smaller that:*

$$1+\min(n_{X'},n_{Y'}).$$

Remark 2. Note that the bound is not tight at all. In our situation the affine bi-linear systems are overdetermined whilst [16] only considered systems with at most as many variables than the number of equations.

Finally, it appears [16] that the matrices occurring during the matrix version of F$_5$ can be made divided into smaller matrices thanks to the bi-linear structure. Let $\dim(R_{d_1,d_2})=\binom{d_1+n_{X'}}{d_1}\binom{d_2+n_{Y'}}{d_2}$. More precisely, the matrices occurring at degree D during the matrix F$_5$ on a bi-linear systems are of size: $\left(\dim(R_{d_1,d_2})-[t_1^{d_1}t_2^{d_2}]\mathrm{HS}(t_1,t_2)\right)\times\dim(R_{d_1,d_2})$ for all (d_1,d_2) such that $d_1+d_2=D, 1\leq d_1,d_2\leq D-1$, where the notation $[t_1^{d_1}t_2^{d_2}]\mathrm{HS}(t_1,t_2)$ stands for the coefficient of the term $t_1^{d_1}t_2^{d_2}$ in the Hilbert bi-serie $\mathrm{HS}(t_1,t_2)$ defined in the appendix.

As pointed out, these results hold for a bi-linear system. For an affine bi-linear, this can be considered as a good (i.e. first order) approximation. The idea is that we have to "bi-homogenize" the affine bi-linear system which corresponds to add some columns. We can then estimate the space/time complexity of computing a Gröbner basis of $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$.

Proposition 5. *Let* $D=\min(n_{X'}+1,n_{Y'}+1)$*. The time complexity of computing a DRL-Gröbner basis* G_{DRL} *of* $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ *is bounded from above by:*

$$\left(\sum_{\substack{d_1+d_2=D\\1\leq d_1,d_2\leq D-1}}\left(\dim(R_{d_1,d_2})-[t_1^{d_1}t_2^{d_2}]\mathrm{HS}(t_1,t_2)\right)^{\omega}\dim(R_{d_1,d_2})\right), \textit{with } \omega, 1\leq\omega\leq 2.$$

The space complexity is bounded by:

$$\left(\sum_{\substack{d_1+d_2=D\\1\leq d_1,d_2\leq D-1}}\left(\dim(R_{d_1,d_2})-[t_1^{d_1}t_2^{d_2}]\mathrm{HS}(t_1,t_2)\right)\dim(R_{d_1,d_2})\right),$$

It is worth to mention that, for the cryptosystems considered in [18], the number of free variables $n_{Y'}$ in $\mathbf{Y}'$ can be rather small (typically 1 or 2 for some challenges). We have then a theoretical explanation of the practical efficiency observed in [18]. In addition, we have a concrete criteria to evaluate the security of future compact McEliece's variants, namely the minimum of the number of variables $n_{X'}$ and $n_{Y'}$ in the blocks $\mathbf{X}'$ and $\mathbf{Y}'$ respectively should be sufficiently "big". This will be further discussed in the last section.

To conclude this section, we mention that the goal of the attack is compute the variety (i.e. set of solutions) $\mathscr{V}$ associated to $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$. As soon as we have a DRL-Gröbner basis G_{DRL} of $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$, the variety can be obtained in $\mathscr{O}\big((\#\mathscr{V})^{\omega}\big)$ thanks to a change of ordering algorithm [17]. We have to be sure that the variety $\mathscr{V}$ has few solutions. In particular, we have to remove parasite solutions (corresponding to $X_i = X_j$ or to $Y_j = 0$). A classical way to do that is to introduce new variables u_{ij} and v_i and add to $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ equations of the form: $u_{ij}\cdot(X_i - X_j) + 1 =$ and $v_i \cdot Y_i + 1 = 0$. In practice, we have not added all theses equations; but only few of them (namely 4 or 5). The reason is that we do not want to add too many new variables. These equations and variables can be added to $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ whilst keeping the affine bi-linear structure. To do so, we have to add the v_i to the block $\mathbf{X}'$, and the variables u_{ij} to the block $\mathbf{Y}'$. So, as we add only few new variables, the complexity of solving $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ with these new constraints is essentially similar to Proposition 5.

5 Application to Key Recovery Attacks of Compact McEliece Variants

The algebraic approach as described in Section 3 had been applied in [18] to two variants of the McEliece cryptosystem [6, 25]. These two systems propose code-based public-key cryptosystems with compact keys by using structured matrices. The BCGO cryptosystem in [6] relies on quasi-cyclic alternant codes whereas the MB cryptosystem in [25] uses quasi-dyadic Goppa codes. The most important fact is that the introduction of structured matrices induces linear relations between the x_i's and the y_j's defining the secret code. We briefly recall how they are built and we refer the reader to [18] for more details.

In both schemes, the public code $\mathscr{C}$ is defined over a field $\mathbb{F}_q = \mathbb{F}_{2^s}$ which is considered as a subfield of $\mathbb{F}_{q^m}$ for a certain integer m.The length n and the dimension k of $\mathscr{C}$ are always of the form $n = n_0\ell$ and $k = k_0\ell$ where ℓ divides $q^m - 1$ and n_0 and k_0 are integers such that $k < n < q^m$. We now give the additional linear equations that link the x_i's and the y_i's in order to describe how the codes are obtained.

BCGO Scheme. Let α be a primitive element of $\mathbb{F}_{q^m}$. Let ℓ and N_0 be such that $q^m - 1 = N_0\ell$ and let β be an element of $\mathbb{F}_{q^m}$ of order ℓ, that is to say $\beta \overset{\text{def}}{=} \alpha^{N_0}$. The public code is an alternant code $\mathscr{A}_r(x,y)$ such that $rm = n - k = (n_0 - k_0)\ell$ and where $x = (x_0,\ldots,x_{n-1})$ and $y = (y_0,\ldots,y_{n-1})$ satisfy for any $b \in \{0,\ldots,n_0-1\}$ and for any $j \in \{0,\ldots,\ell-1\}$ the following linear equations [18]:

$$\begin{cases} x_{b\ell+j} = x_{b\ell}\beta^{j} \\ \\ y_{b\ell+j} = y_{b\ell}\beta^{je} \end{cases} \tag{9}$$

where e is an integer secretly picked in $\{0,\ldots,\ell-1\}$. We are able to simplify the description of the system $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$ by setting up the unknown X_b for $x_{b\ell}$ and the unknown Y_b for $y_{b\ell}$. We obtain the following algebraic system in which we assume that e is known:

Proposition 6 ([18]). *Let $G = (g_{i,j})$ be the $k \times n$ public generator matrix with $k = k_0\ell$ and $n = n_0\ell$. For any $0 \le w \le r-1$ and any $0 \le i \le k-1$, the unknowns $X_0,\ldots,X_{n_0-1}$ and $Y_0,\ldots,Y_{n_0-1}$ should satisfy:*

$$\sum_{b=0}^{n_0-1} g'_{i,b,w} Y_b X_b^{w} = 0 \qquad \textit{where } g'_{i,b,w} \overset{def}{=} \sum_{j=0}^{\ell-1} g_{i,b\ell+j}\beta^{j(e+w)}. \tag{10}$$

Furthermore, one X_i can be set to any arbitrary value (say $X_0 = 0$) as well as one Y_i can be set to any arbitrary nonzero value (say $Y_0 = 1$). Finally, The system (10) has $(n_0 - 1)$ unknowns Y_i and $(n_0 - 1)$ unknowns X_i. It has k_0 linear equations involving only the Y_i's and $(r-1)k/\ell = (r-1)k_0$ polynomial equations involving the monomials $Y_iX_i^w$ with $w > 0$.

We have then:

Corollary 1. *The system* (10) *has $n_{Y'} = n_0 - k_0 - 1$ free variables in the Y_b's.*

MB Scheme. The public code defined by the (public) generator matrix G can be seen as an alternant code $\mathscr{A}_\ell(x,y)$ (that is to say $r=\ell$) where for any $0 \leq b \leq n_0-1$ and $0 \leq i \leq \ell-1$, we have the following linear equations [18]:

$$\begin{cases} y_{b\ell+i} = y_{b\ell} \\ \\ x_{b\ell+i} = x_{b\ell} + \sum_{j=0}^{\log_2(\ell-1)} \eta_j(i)(x_{2^j}+x_0) \end{cases} \tag{11}$$

where $\sum_{j=0}^{\log_2(\ell-1)} \eta_j(i)2^j$ with $\eta_j(i) \in \{0,1\}$ is the binary decomposition of i. This description enables to simplify the unknowns involved in $\mathsf{McE}_{k,n,r}(\mathbf{X},\mathbf{Y})$ to $Y_{b\ell}$, $X_{b\ell}$ with $b \in \{0,\dots,n_0-1\}$ and to the unknowns X_{2^j} with $j \in \{0,\dots,\log_2(\ell-1)\}$ We then obtain the following algebraic system:

Proposition 7 ([18]). *Let $G=(g_{i,j})$ be the $k \times n$ public generator matrix with $k=k_0\ell$ and $n=n_0\ell$. For any w, i such that $0 \leq w \leq \ell-1$ and $0 \leq i \leq k-1$, we have:*

$$\sum_{b=0}^{n_0-1} Y_{b\ell} \sum_{j=0}^{\ell-1} g_{i,b\ell+j} \left(X_{b\ell} + \sum_{j=0}^{\log_2(\ell-1)} \eta_j(j)(X_{2^j}+X_0) \right)^w = 0 \tag{12}$$

Furthermore, two X_i's can be set to any (different) arbitrary values (say $X_0=0$ and $X_1=1$) as well as one Y_i can be set to any arbitrary nonzero value (say $Y_0=1$). Finally, The system (12) has n_0-1 unknowns Y_i and $n_0-2+\log_2(\ell)$ unknowns X_i. Furthermore, it has n_0-m linear equations involving only the Y_i's, and $(\ell-1)\ell(n_0-m)$ polynomial equations involving the monomials $Y_iX_i^w$ with $w>0$.

It holds then:

Corollary 2. *The system* (12) *has $n_{Y'}=m-1$ free variables in the Y_b's.*

6 Comparison of Theoretical complexity with Experimental Results

In the table below, we present the experimental results obtained in [18] for BCGO and MB schemes. For the sack of comparaison, we include a bound on theoretical complexity of computing a Gröbner bases of $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$:

$$T_{\text{theo}} \approx \left(\sum_{\substack{d_1+d_2=D \\ 1 \leq d_1,d_2 \leq D-1}} \Big(\dim(R_{d_1,d_2}) - [t_1^{d_1}t_2^{d_2}]\mathrm{HS}(t_1,t_2) \Big) \dim(R_{d_1,d_2}) \right), \tag{13}$$

as obtained in Section 4. Regarding the linear algebra, this is a bit optimistic. However, as already pointed our, we have been also rather pessimistic regarding others parameters. For instance, we are not using the fact that the systems are overdetermined, and we have also only considered a sub-system of $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$.

All in all, this bound permits a give a reasonable picture of the hardness of solving $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$. It is of course not sufficient to set parameters, but sufficient to discard many weak compact variants of McEliece.

Table 1. Cryptanalysis results for [6] ($m = 2$)

Challenge	q	ℓ	n_0	$n_{Y'}$	Security [6]	$n_{X'}$	Equations	Time (Operations, Memory)	T_{theo}
A_{16}	2^8	51	9	3	80	8	510	0.06 sec ($2^{18.9}$ op, 115 Meg)	2^{17}
B_{16}	2^8	51	10	3	90	9	612	0.03 sec ($2^{17.1}$ op, 116 Meg)	2^{18}
C_{16}	2^8	51	12	3	100	11	816	0.05 sec ($2^{16.2}$ op, 116 Meg)	2^{20}
D_{16}	2^8	51	15	4	120	14	1275	0.02 sec ($2^{14.7}$ op, 113 Meg)	2^{26}
A_{20}	2^{10}	75	6	2	80	5	337	0.05 sec ($2^{15.8}$ op, 115 Meg)	2^{10}
B_{20}	2^{10}	93	6	2	90	5	418	0.05 sec ($2^{17.1}$ op, 115 Meg)	2^{10}
C_{20}	2^{10}	93	8	2	110	7	697	0.02 sec ($2^{14.5}$ op, 115 Meg)	2^{11}
QC$_{600}$	2^8	255	15	3	600	14	6820	0.08 sec ($2^{16.6}$ op, 116 Meg)	2^{21}

Table 2. Cryptanalysis results for [25].

Challenge	q	$n_{Y'}$	ℓ	n_0	Security	$n_{X'}$	Equations	Time (Operations, Memory)	T_{theo}
Table 2	2^2	7	64	56	128	59	193,584	1,776.3 sec ($2^{34.2}$ op, 360 Meg)	2^{65}
Table 2	2^4	3	64	32	128	36	112,924	0.50 sec ($2^{22.1}$ op, 118 Meg)	2^{29}
Table 2	2^8	1	64	12	128	16	40,330	0.03 sec ($2^{16.7}$ op, 35 Meg)	2^8
Table 3	2^8	1	64	10	102	14	32,264	0.03 sec ($2^{15.9}$ op, 113 Meg)	2^8
Table 3	2^8	1	128	6	136	11	65,028	0.02 sec ($2^{15.4}$ op, 113 Meg)	2^7
Table 3	2^8	1	256	4	168	10	130,562	0.11 sec ($2^{19.2}$ op, 113 Meg)	2^7
Table 5	2^8	1	128	4	80	9	32,514	0.06 sec ($2^{17.7}$ op, 35 Meg)	2^6
Table 5	2^8	1	128	5	112	10	48,771	0.02 sec ($2^{14.5}$ op, 35 Meg)	2^7
Table 5	2^8	1	128	6	128	11	65,028	0.01 sec ($2^{16.6}$ op, 35 Meg)	2^7
Table 5	2^8	1	256	5	192	11	195,843	0.05 sec ($2^{17.5}$ op, 35 Meg)	2^7
Table 5	2^8	1	256	6	256	12	261,124	0.06 sec ($2^{17.8}$ op, 35 Meg)	2^7
Dyadic$_{256}$	2^4	3	128	32	256	37	455,196	7.1 sec ($2^{26.1}$ op, 131 Meg)	2^{29}
Dyadic$_{512}$	2^8	1	512	6	512	13	1,046,532	0.15 sec ($2^{19.7}$ op, 38 Meg)	2^8

We briefly discussed of the theoretical complexity obtained for the first row of the second column. As explained, we have used the formula (13). We have computed the coefficient $[t_1^{d_1}t_2^{d_2}]\mathrm{HS}(t_1,t_2)$ by using the explicit formula of $\mathrm{HS}(t_1,t_2)$ provided in the appendix using the explicit values of $n_{X'} = 59$ and $n_{Y'}$=7, and assuming that the system is square; in that case the degree of regularity is 8. For this parameter, the sub-system $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ has actually 288 equations (of degree $2,3$ and 5). Hence, it is interesting to compute [2,4,3,5] the degree of regularity of a semi-regular system of the same size: we found a regularity of 11 leading to a cost of $2^{85.2}$ for the Gröbner basis computation (using (6), with $\omega = 2$). It is expected that a new results of the degree of regularity of generic overdetermined bi-linear systems would lead to tighter bounds.

As a conclusion, one can see that the theoretical bound (13) provides a reasonable explanation regarding the efficiency of the attack presented in [18]. In particular, it is important to remark that the hardness of the attack seems related to $d = \min(n_X', n_Y')$. The complexity of the attack clearly increases with this quantity. For the design of future compact variants of McEliece, this d should be then not too small. Regarding the current state of the art, it is difficult to provide an exact value. Very roughly speaking, $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ can be considered as hard as solving a random (overdetermined) algebraic system with $d = \min(n_{X'}, n_{Y'})$ equations over a big field. With this in mind, we can say that any system with $d \leq 20$ should be within the scope of an algebraic attack.

Note that another phenomena, which remains to be treated, can occur. In the particular case of binary dyadic codes, the Gröbner basis of $\mathsf{BiMcE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ can be easily computed, but the variety associated is too big. This is due to the fact that the Gröbner basis is “trivial" (reduced to one equation) and not provides then enough information. This is typically due to the fact that we have used only a sub-set of the equations $\big($of bi-degree $(2^j,1)$). So, the open question is how we can use cleverly all the equations of $\mathsf{McE}_{k,n,r}(\mathbf{X}',\mathbf{Y}')$ in the binary case.

References

1. M. Baldi and G. F. Chiaraluce. Cryptanalysis of a new instance of McEliece cryptosystem based on QC-LDPC codes. In *IEEE International Symposium on Information Theory*, pages 2591–2595, Nice, France, March 2007.
2. Magali Bardet. *Étude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie*. PhD thesis, Université de Paris VI, 2004.
3. Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. Complexity study of Gröbner basis computation. Technical report, INRIA, 2002. http://www.inria.fr/rrrt/rr-5049.html.
4. Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations. In *Proc. International Conference on Polynomial System Solving (ICPSS)*, pages 71–75, 2004.
5. Magali Bardet, Jean-Charles Faugère, Bruno Salvy, and Bo-Yin Yang. Asymptotic behaviour of the degree of regularity of semi-regular polynomial systems. In *Proc. of MEGA 2005, Eighth International Symposium on Effective Methods in Algebraic Geometry*, 2005.
6. T. P. Berger, P.L. Cayrel, P. Gaborit, and A. Otmani. Reducing key length of the McEliece cryptosystem. In Bart Preneel, editor, *Progress in Cryptology - Second International Conference on Cryptology in Africa (AFRICACRYPT 2009)*, volume 5580 of *Lecture Notes in Computer Science*, pages 77–97, Gammarth, Tunisia, June 21-25 2009.
7. E. Berlekamp, R. McEliece, and H. van Tilborg. On the inherent intractability of certain coding problems. *Information Theory, IEEE Transactions on*, 24(3):384–386, May 1978.
8. E. R. Berlekamp. Factoring polynomials over finite fields. In E. R. Berlekamp, editor, *Algebraic Coding Theory*, chapter 6. McGraw-Hill, 1968.
9. D. J. Bernstein, T. Lange, and C. Peters. Attacking and defending the McEliece cryptosystem. In *PQCrypto*, volume 5299 of *LNCS*, pages 31–46, 2008.
10. B. Biswas and N. Sendrier. McEliece cryptosystem implementation: Theory and practice. In *PQCrypto*, volume 5299 of *LNCS*, pages 47–62, 2008.
11. Bruno Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. PhD thesis, University of Innsbruck, 1965.
12. A. Canteaut and F. Chabaud. A new algorithm for finding minimum-weight words in a linear code: Application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511. *IEEE Transactions on Information Theory*, 44(1):367–378, 1998.
13. D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, Varieties, and algorithms: an Introduction to Computational Algebraic Geometry and Commutative Algebra.* Undergraduate Texts in Mathematics, Springer-Verlag, New York., 2001.
14. J.-C. Faugère. A new efficient algorithm for computing gröbner bases (f4). *Journal of Pure and Applied Algebra*, 139(1-3):61–88, 1999.
15. J.-C. Faugère. A new efficient algorithm for computing gröbner bases without reduction to zero : F5. In *ISSAC'02*, pages 75–83. ACM press, 2002.
16. Jean-Charles Faugère, Mohab Safey El Din, and Pierre-Jean Spaenlehauer. Gröbner bases of bihomogeneous ideals generated by polynomials of bidegree (1,1): Algorithms and complexity. *CoRR*, abs/1001.4004, 2010.
17. Jean-Charles Faugère, Patrizia M. Gianni, Daniel Lazard, and Teo Mora. Efficient computation of zero-dimensional gröbner bases by change of ordering. *J. Symb. Comput.*, 16(4):329–344, 1993.
18. Jean-Charles Faugère, Ayoub Otmani, Ludovic Perret, and Jean-Pierre Tillich. Algebraic cryptanalysis of McEliece variants with compact keys. In *Proceedings of Eurocrypt 2010*. Springer Verlag, 2010. to appear.
19. P. Gaborit. Shorter keys for code based cryptography. In *Proceedings of the 2005 International Workshop on Coding and Cryptography (WCC 2005)*, pages 81–91, Bergen, Norway, March 2005.
20. P. J. Lee and E. F. Brickell. An observation on the security of McEliece's public-key cryptosystem. In *Advances in Cryptology - EUROCRYPT'88*, volume 330/1988 of *Lecture Notes in Computer Science*, pages 275–280. Springer, 1988.
21. J. S. Leon. A probabilistic algorithm for computing minimum weights of large error-correcting codes. *IEEE Transactions on Information Theory*, 34(5):1354–1359, 1988.
22. P. Loidreau and N. Sendrier. Weak keys in the mceliece public-key cryptosystem. *IEEE Transactions on Information Theory*, 47(3):1207–1211, 2001.
23. F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North–Holland, Amsterdam, fifth edition, 1986.

24. R. J. McEliece. *A Public-Key System Based on Algebraic Coding Theory*, pages 114–116. Jet Propulsion Lab, 1978. DSN Progress Report 44.
25. R. Misoczki and P. S. L. M. Barreto. Compact McEliece keys from Goppa codes. In *Selected Areas in Cryptography (SAC 2009)*, Calgary, Canada, August 13-14 2009.
26. A. Otmani, J.P. Tillich, and L. Dallot. Cryptanalysis of McEliece cryptosystem based on quasi-cyclic ldpc codes. In *Proceedings of First International Conference on Symbolic Computation and Cryptography*, pages 69–81, Beijing, China, April 28-30 2008. LMIB Beihang University.
27. N. Patterson. The algebraic decoding of Goppa codes. *IEEE Transactions on Information Theory*, 21(2):203–207, 1975.
28. N. Sendrier. Finding the permutation between equivalent linear codes: The support splitting algorithm. *IEEE Transactions on Information Theory*, 46(4):1193–1203, 2000.
29. J. Stern. A method for finding codewords of small weight. In G. D. Cohen and J. Wolfmann, editors, *Coding Theory and Applications*, volume 388 of *Lecture Notes in Computer Science*, pages 106–113. Springer, 1988.

A Hilbert Bi-Series

We say that an ideal is *bihomogeneous* if there exists a set of bihomogeneous generators. The vector space of bihomogeneous polynomials of bi-degree (α,β) in a polynomial ring R will be denoted by $R_{\alpha,\beta}$. If $\mathscr{I}$ is a bihomogeneous ideal, then $I_{\alpha,\beta}$ will denote the vector space $I \cap R_{\alpha,\beta}$.

Definition 2 ([16]). *Let $\mathscr{I}$ be a bihomogeneous ideal of R. The Hilbert bi-series is defined by*

$$\mathrm{HS}_{\mathscr{I}}(t_1,t_2) = \sum_{(\alpha,\beta)\in\mathbb{N}^2} \dim(R_{\alpha,\beta}/I_{\alpha,\beta}) t_1^{\alpha} t_2^{\beta}.$$

For bi-regular bilinear systems, [16] provide an explicit form of the bi-series.

Theorem 2. *Let $f_1,\dots,f_m \in R$ be a bi-regular bilinear sequence, with $m \le n_{X'} + n_{Y'}$. Then*

$$\mathrm{HS}_{I_m}(t_1,t_2) = \frac{(1-t_1t_2)^m + N_{n_{X'}+1}(t_1,t_2) + N_{n_{Y'}+1}(t_1,t_2)}{(1-t_1)^{n_{X'}+1}(1-t_2)^{n_{Y'}+1}},$$

where

$$N_n(t_1,t_2) = t_1t_2(1-t_2)^n \sum_{\ell=1}^{m-n} (1-t_1t_2)^{m-n-\ell} \left[1-(1-t_1)^{\ell} \sum_{k=1}^{n} t_1^{n-k} \binom{\ell+n-k-1}{n-k}\right].$$