ISSN 1996-3343

Asian Journal of Applied Sciences

http://knowledgiascientific.com

Knowledgia SCIENTIFIC
A Place to Publish Outstanding Research

Asian Journal of Applied Sciences 8 (1): 37-45, 2015
ISSN 1996-3343 / DOI: 10.3923/ajaps.2015.37.45
© 2015 Knowledgia Review, Malaysia

Blind Reconstruction of RS Code

Lu Ouxin, Gan Lu and Liao Hongshu
School of Electronic Engineering, University of Electronic Science and Technology of China, China

Corresponding Author: Lu Ouxin, School of Electronic Engineering, University of Electronic Science and Technology of China, China

ABSTRACT

In this study, the Gröbner bases of modules are employed as a tool to analyze block length of RS code. A coordinate permutation of a RS code, which is defined over GF(p^m) in GF(p), the polynomial vector form of the result can be considered as a sub module. Based on the algebraic structure of quasi-cyclic codes, the Gröbner bases of such a sub-module shed light on block length and dimension of the original code. Thus, block length of the code can be estimated. Furthermore, the primitive polynomial and generator polynomial of the code are reconstructed through Galois Field Fourier Transform (GFFT) technique.

Key words: Blind estimation, RS code, gröbner bases, submodule, GFFT

INTRODUCTION

In a context of non-cooperative communication or reverse-engineering of a communication system, it is of great interest to recognize the parameters, such as block length and parity check matrix of the codes in the intercepted bit stream. Due to the lack of prior information, the error bits exists in the bit stream. The problem of recognizing and reconstruction can be very tough in some cases.

Till now, the publications cope with this problem are scarce. Some of them focuses on identification and reconstruction of convolution codes (Filiol, 1997; Rice, 1995). As to the block codes, Valembois (2001) has formalized the problem of finding the nearest (for the Hamming distance) block code from bit stream and proved it is NP-complete. In other studies dealing with the problem, Chabot (2007) and Cluzeau (2006) are restricted to find the parity check matrix and may have limitations in applications, such as algorithm works efficiently only when the weight of the dual code is very low. Furthermore, all of these solutions considered only the case of binary codes.

Reed-Solomon (RS) code, a class of non-binary cyclic block codes, is a subfamily of the cyclic BCH codes (Shu *et al.*, 2005). The RS code is optimal in the sense that the minimum distance has the maximum value possible for a linear code of size (n, k) which is known as the Singleton bound (Moreira and Farrell, 2006). It is widely used in mass storage systems, data transmission (such as CCSDS, DVB), as well as space transmission.

As an important subfamily of the cyclic BCH codes, its cyclic property in extension field GF(p^m) ensures that RS code is quasi-cyclic in GF(p). The most useful RS codes in practice are those defined over GF(2^m) (Ling and Sole, 2001). Also, this is the kind of RS codes which is going to be used for demonstration.

In this study, a new approach, which based on Gröbner bases of modules, is adopted to fully exploit the quasi-cyclic property of RS code in GF(2). After a coordinate permutation of a RS code, the polynomial vector form of the result can be treated as a submodule. Due to the algebraic structure of the original code, the Gröbner bases of the submodule have some distinctions which

differentiate itself from the submodules of those incorrectly-estimated codes. Thus, the code can be detected. Through the triangularization of the Gröbner bases of the submodule, if there is no error bit, the dimension of the original code can be calculated at the same time according to the result of Lally and Fitzpatrick (2001). A further step is that GFFT is being used to reconstruct the primitive polynomial and generator polynomial of the original code.

BASIC STRUCTURES AND TOOLS

Modules and submodules: Among the structures in abstract algebra, a module is definitely a versatile one, whose concept derives from the attempt to conduct classical linear algebra with an arbitrary ring, instead of the traditional field (Hartley and Hawkes, 1970). However, this kind of generalization from vector spaces to module spaces at the same time entails a sacrifice because the scalars do not necessarily have inverses. For formal definition, a module has been presented by Shu *et al.* (2005).

In a submodule of an R-module, M will be a subset N of M such that the operations of M, when restricted to N, make N into an R-module. For example; let M be an R-module, $\emptyset \neq U \subset M$ is called a submodule of M if (UM1) $(U, +) < (M, +)$ and (UM2) $\alpha \in R$, $u \in U \Rightarrow \alpha u \in U$, that is, $RU \subset U$.

Gröbner bases of submodules: A Gröbner basis is a specific generating set of an ideal or submodule over a polynomial ring. It is not minimal in general but has extremely nice properties. It is reasonably easy to extract information about the ideal or submodule from a Gröbner basis. Here, the definition of Gröbner bases of submodules is given as a set of non-zero vectors $G = \{g_1, \ldots, g_t\}$ contained in the submodule M is called a Gröbner basis for M if $f \in M$, there exists $i \in \{1, \ldots, t\}$ such that $Im(g_i)$ divide $Im(f)$. The set G is called a Gröbner basis for the submodule (G). where, Im(f) is the leading monomial of f.

Gröbner basis can be obtained from Adams and Loustaunau (1994) and Buchberger and Winkler (1998).

Quasi-cyclic codes and Gröbner bases: Assuming C is a quasi-cyclic code whose index is l, code length is $n = l \times m$. It is obvious that a coordinate permutation with a codeword c of C can be represented in a polynomial vector form of $c = (c_0(x), c_1(x), \ldots c_{l-1}(x))$, will be considered as element of R-submodule space of R^l, where $R = F(x)/(x^m-1)$. The pre-image $\tilde{C}$ of C in $F[x]^l$ is accordingly a F[x]-submodule embracing $\tilde{K} = \langle (x^m - 1)e_i,\ i = 1, 2, \ldots, l-1 \rangle$. Note that e_i is the standard basis vector with 1 in position i and 0 elsewhere. It can be proved that every submodule $\tilde{C}$ of $F[x]^l$ containing $\tilde{K}$ has a reduced Gröbner basis of a specific form (Hartley and Hawkes, 1970). Based on the connection between C and $\tilde{C}$, a natural homomorphism, the relation can be expressed as: $\varphi: F[x]^l \to R^l, (c_0, c_1, \cdots c_{l-1}) \mapsto (c_0 + (x^m+1), c_1 + (x^m+1), \cdots c_{l-1} + (x^m+1))$ the quasi-cyclic code C has a specific generating set which is uniquely defined if the term order is given. Furthermore, the corollary is given directly as follows the dimension of the code C with Gröbner bases generating set $\{\varphi(g_i), i = 0, 1, \ldots, l-1\}$ is given by:

$$lm - \sum_{i=0}^{l-1} \partial g_{ii} = \sum_{i=0}^{l-1} (m - \partial g_{ii})$$

(Lally and Fitzpatrick, 2001).
where, ∂ is degree of a polynomial.

Galois Field Fourier Transform (GFFT): Let $a(x) = a_{n-1}x^{n-1}+a_{n-2}x^{n-2}+a_{n-3}x^{n-3}+\ldots+a_0$ be a polynomial over GF(q), where $a_i \epsilon GF(q)$, then the Mattson-Solomon (MS) polynomial of a(x) is defined as:

$$A(z) = A_{n-1}z^{n-1} + A_{n-2}z^{n-2} + A_{n-3}z^{n-3} + \ldots + A_0 = \sum_{j=0}^{n-1} A_j z^j$$

Where:

$$A_j = \sum_{i=0}^{n-1} a_i \cdot \alpha^{ji},\ j = 0, 1, 2, \ldots, n-1$$

α is a primitive nth root of unity over $GF(q^m)$. $A = (A_{n-1}, A_{n-2}, A_{n-3}, \ldots A_0)$ is Galois Field Fourier Transform (GFFT) of $a = (a_{n-1}, a_{n-2}, a_{n-3}, \ldots a_0)$ (Lally and Fitzpatrick, 2001) which can be defined in matrix form as:

$$\begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ \vdots \\ A_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ 1 & \alpha^2 & (\alpha^2)^2 & \cdots & (\alpha^2)^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha^{n-1} & (\alpha^{n-1})^2 & \cdots & (\alpha^{n-1})^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix}$$

Let the Vandermonde matrix on the right side of equation be matrix V, the equation above can be written as:

$$A = Va$$

Thus, according to the definition of RS code, the GFFT result of the polynomial form of its code word contains a series of successive zeros.

PROPOSED ALGORITHM

Coordinate permutation of original codes: In order to make use of Gröbner bases tool, the first step of the algorithm is to preprocess the intercepted code words to transform them into a specific submodule spaces. The coordinate permutation method is utilized within a codeword as explained below. Given a codeword c of a quasi-cyclic code C, whose index is l, code length is $n = l \times m$, c can be represented as:

$$c = (c_{0,0}, c_{0,1}, \ldots, c_{0,I-1}, c_{1,0}, \ldots, c_{1,I-1}, \ldots, c_{m,0}, \ldots, c_{m,I-1})$$

Through a coordinate permutation it can be expressed as:

$$\begin{bmatrix} c_{0,0} & c_{0,1} & \cdots & c_{0,l-1} \\ c_{1,0} & c_{1,1} & \cdots & c_{0,l-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m-1,0} & c_{m-1,1} & \cdots & c_{m-1,l-1} \end{bmatrix} \Rightarrow \quad \phi(c) = (c_0(x), c_1(x), \cdots c_{l-1}(x)) \in R^1$$

Where:

$$c_j(x) = \sum_{i=0}^{m-1} c_{ij} x^i \in R,\ R = \frac{F(x)}{(x^m - 1)}$$

Calculation of the Gröbner bases of derived submodule spaces: The process of computing a Gröbner is actually a generalization of three familiar methods: Gaussian elimination for solving linear systems of equations, the Euclidean algorithm for computing the greatest common divisor of two univariate polynomials and the Simplex Algorithm for linear programming. Buchberger and Winkler (1998) were first to give an algorithm for computing the Gröbner bases in dissertation in 1965. Currently, there are several Computer Algebra Systems (CAS) in hand to compute the Gröbner bases and the primary concern of this study is applying the Gröbner bases to estimate the parameters of RS codes but not the computation algorithm itself. One of the CAS, namely Macaulay2, is employed as a tool to compute the Gröbner bases of submodule spaces (http://math.uiuc.edu/Macaulay2/).

Estimating the parameters of the original code: According to the properties of the Gröbner bases of Quasi-cyclic codes as mentioned in above, whether or not a sequence of intercepted code-words belongs to a quasi-cyclic code can be detected by the properties of Gröbner bases of submodule spaces. Rather, the Gröbner bases of submodule spaces of incorrectly estimated codewords which can be considered as randomly generated bit sequence, is an identity matrix as:

$$I_{l\times 1} = \{e_i,\ i = 0, 1,..., l-1\}$$

or is very close to $I_{l\times 1}$. Here, r_a which is the proportion of e_i in the derived Gröbner base matrix, is defined to indicate its similarity to identity matrix:

$$r_a = \frac{num(e_i)}{l}$$

where, $num(e_i)$ is the number of e_i in the matrix.

Thus, if $r_a = 1$, the matrix is completely an identity matrix. We choose n_{est} whose corresponding $r_a = \min\{r_{a1}, r_{a2},\ldots, r_{aj}\}$ as an estimation of block length n.

Moreover, if the intercepted code words contains no error bit, its dimension can be estimated directly that is information length k, directly by the corollary (estimating the parametesr of original code).

Subsequently, based on the block length n_{est} of the original code which has been estimated above, GFFT is used to reconstruct the primitive polynomial and generator polynomial according to the characteristics. Let A_i be the GFFT result of the code word according to i primitive polynomial, then, p_i over $GF(2^m)$:

$$A_i = (A_{i1}, A_{i2}, A_{i3},\ldots, A_{in_{est}})$$

If $A_{i1} = A_{i2} = \cdots = A_{ik_{est}} = 0$, the estimated primitive polynomial $p_{est} = p_i$, information length $k_{est} = k'$. Accordingly, the generator polynomial g can be reconstructed.

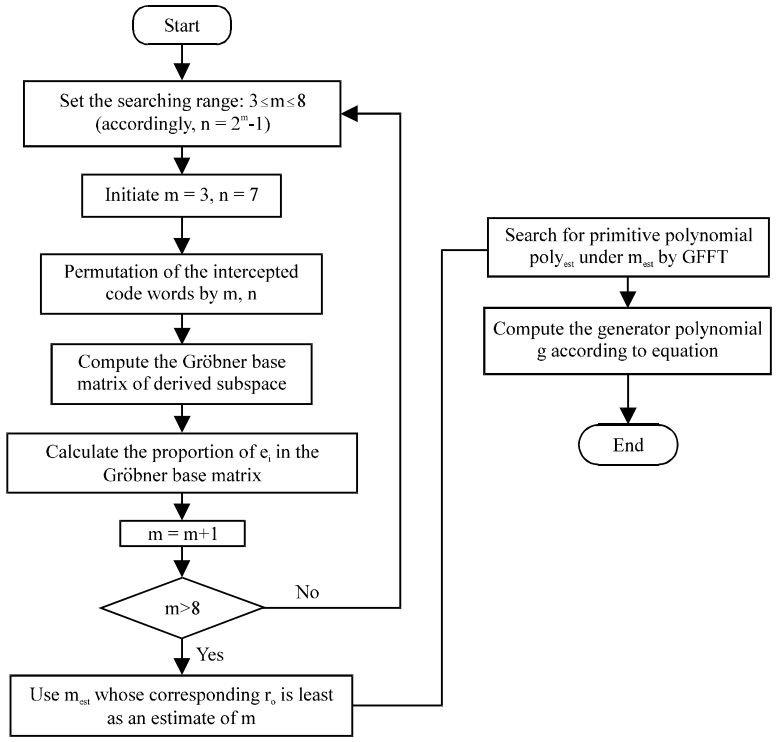

Fig. 1: Flow chart of the estimation of RS code

The algorithm is showed in a flow chart (Fig. 1).

METHODOLOGY

RS(7, 3) code, whose primitive polynomial is x^3+x^2+1, is used for the demonstration of the reconstruction process. As a cyclic code in GF(2^m), RS code is actually a quasi-cyclic code in GF(2). Like many other blind estimation methods, the assumption can be made in order to synchronize the intercepted code words sequence. Once synchronized the first step is to search for code length n. When n_{est} is assumed, its correspondent code width m_{est} can be identified.

After the coordinate permutation of the original code words, whether or not the code words belong to a quasi-cyclic code can be judged through the number of e_i contained in the Gröbner bases of derived submodule space. For comparison, r_a, the proportion of e_i, is defined in order to measure how close the derived Gröbner base matrix is to be the identity matrix.

Once the search has been completed, the n'_{est} with the Gröbner base matrix which contains least proportion of e_i is chosen as the estimation of code length n as follows:

$$\begin{vmatrix} x^6+x^5+x^4+x^3+x^2+x+1 & x^5+x+1 & x^4+x^2+x+1 \\ 0 & x^3+x^2+1 & 0 \\ 0 & 0 & x^3+x^2+1 \end{vmatrix}$$

where, $r_a = 0$.

If the code length n_{est} is incorrect or there are too much error bits in the intercepted words sequence, the Gröbner base matrix is an identity matrix as follows:

$$I_{m\times m}=\begin{vmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 1\end{vmatrix}$$

where, $r_a = 1$.

One step further, if the intercepted code words sequence contains no error bit. Estimating the parameters of the original code can be used to estimate the information length k from the triangularized form of Gröbner base matrix.

According to estimating the parameters of the original code:

$$k=\sum_{i=0}^{l-1}(m-\partial g_{ii})=(7-6)+(7-3)+(7-3)=9$$

Accordingly:

$$r=\frac{k}{m_{est}\times n_{est}}=\frac{9}{3\times 7}=\frac{3}{7}$$

thus, the original code is an RS(7, 3) code.

However, error bits exist. So, here GFFT is used to estimate the primitive polynomial and generator polynomial of the original RS code. Under the estimated code length n'_{est}, the possible primitive polynomials are determined. Each of them is used to do GFFT in $GF(2^m)$. According to successive zeros of the results, the primitive polynomials and generator polynomials can be estimated.

If the primitive polynomial is correctly assumed as x^3+x^2+1, the result is shown as follows:

$$(5\quad 7\quad 3\quad 0\quad 0\quad 0\quad 0)$$

$$(3\quad 5\quad 1\quad 0\quad 0\quad 0\quad 0)$$

$$(4\quad 6\quad 3\quad 0\quad 0\quad 0\quad 0)$$

From the number of the successive zeros in the results of GFFT, the information length of the original code can be estimated as k = 3. Thus, the original code is RS(7, 3) code, whose primitive polynomial is x^3+x^2+1, generator polynomial is $g(x) = x^4+4x^3+5x^2+1$.

RESULTS

Some RS codes with comparatively larger block length and different number of code words are used in Monte Carlo simulation in order to verify the robustness of the reconstruction algorithm. They are RS(127, 63) with 7 code words, RS(255, 127) with 8 code words and RS(255, 127) with 25 code words and the corresponding performances of the estimating method are shown in Fig. 2-4, respectively as follows:

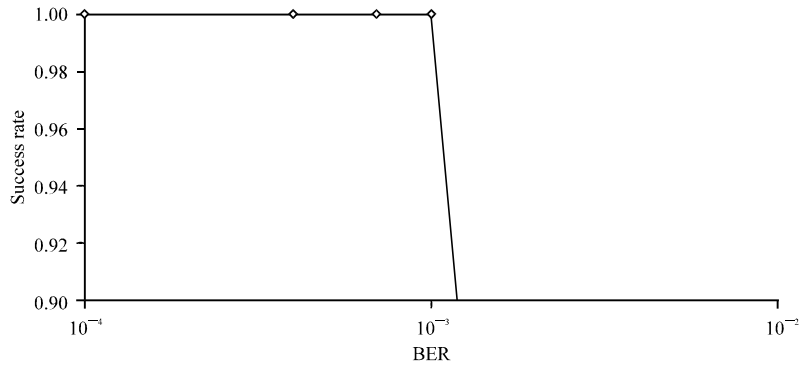

Fig. 2: Reed-Solomon (RS) (127, 63) with 7 code words

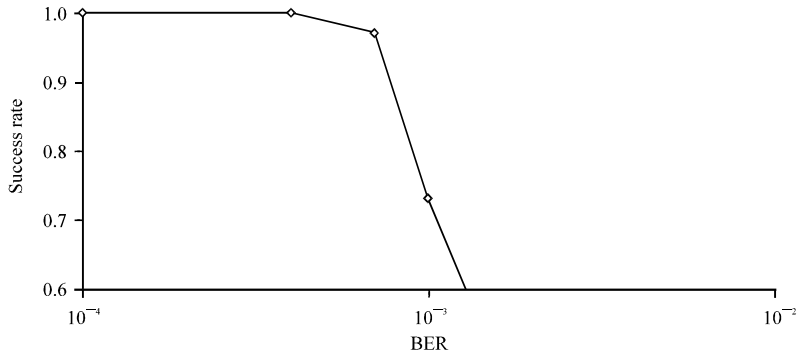

Fig. 3: Reed-Solomon (RS) (255, 127) with 8 code words

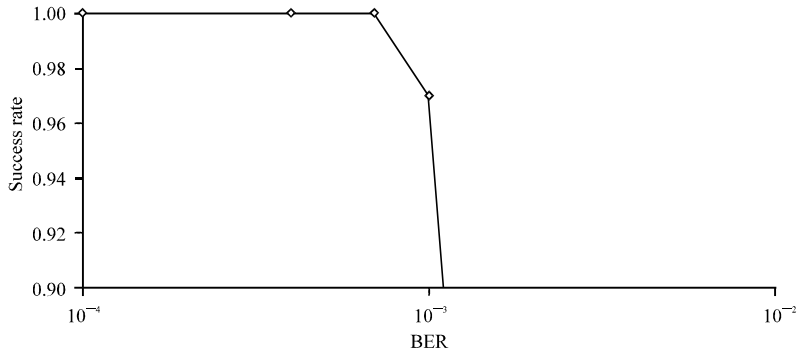

Fig. 4: Reed-Solomon (RS) (255, 127) with 25 code words

Figure 2 shows that when block length is 127 and the number of intercepted code words is 7, the performance of the proposed method is satisfying that is, even when the bit error rate is 0.001, the successful reconstruction rate is 100%. While in Fig. 3, when the bit error rate is still 0.001, the successful reconstruction rate falls down to 0.74 approximately. This sharp difference illustrates that the block length of the RS code is a factor which has an impact on reconstruction performance.

Furthermore, through the comparison between Fig. 3-4, it has been suggested that the number of the code words used in the reconstruction process can affect the result even when other parameters stay the same.

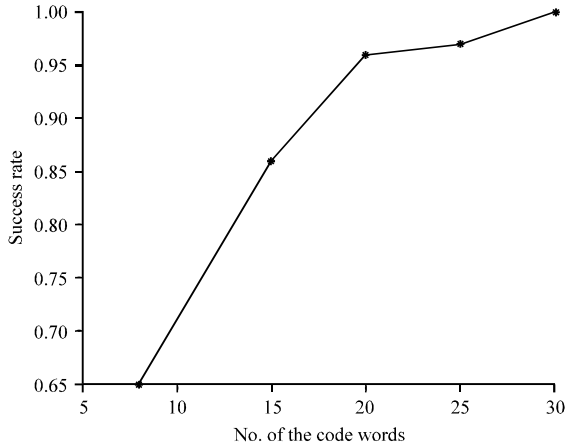

Fig. 5: Relationship between number of codes and performance

When the number of error bits in the intercepted code words sequence is considerable, the number of code words used plays an important role. In a certain scope, more data is used, more likely the estimating results are correct. The output of Monte Carlo simulation reflects the relationship in Fig. 5 as follows.

CONCLUSION

In this study, a reconstruction method of Reed-Solomon codes is given. The method can be mainly divided into two steps. In the first step, the tool of Gröbner base is used to efficiently identify the block length. This could be generalized to all the codes with quasi-cyclicity. In the second step, Galois Field Fourier Transform (GFFT) is employed to reconstruct the primitive polynomial and generator polynomial. Simulation results show that the method is both valid and robust.

ACKNOWLEDGMENTS

This study was supported by the NSAF of China (Grant No. 11176005).

REFERENCES

Adams, W.W. and P. Loustaunau, 1994. An Introduction to Grobner Bases. American Mathematical Society, USA., ISBN-13: 9780821872161, Pages: 289.

Buchberger, B. and F. Winkler, 1998. Grobner Bases and Applications. Cambridge University Press, Cambridge, UK., ISBN-13: 9780521632980, Pages: 552.

Chabot, C., 2007. Recognition of a code in a noisy environment. Proceedings of the IEEE International Symposium on Information Theory, June 24-29, 2007, Nice, France, pp: 2210-2215.

Cluzeau, M., 2006. Block code reconstruction using iterative decoding techniques. Proceedings of the IEEE International Symposium on Information Theory, July 14-19, 2006, Seattle, WA., pp: 2269-2273.

Filiol, E., 1997. Reconstruction of Convolutional Encoders Over GF (q). In: Crytography and Coding, Darnell, M. (Ed.). Springer, Berlin, Heidelberg, ISBN: 978-3-540-63927-5, pp: 101-109.

Hartley, B. and T.O. Hawkes, 1970. Rings, Modules and Linear Algebra. Chapman and Hall Ltd., UK., pp: 69-77.

Lally, K. and P. Fitzpatrick, 2001. Algebraic structure of quasicyclic codes. Discrete Applied Math., 111: 157-175.

Ling, S. and P. Sole, 2001. On the algebraic structure of quasi-cyclic codes. I. Finite fields. IEEE Trans. Informat. Theory, 47: 2751-2760.

Moreira, J.C. and P.G. Farrell, 2006. Essentials of Error-Control Coding. John Wiley and Sons Ltd., England, ISBN-10: 047002920X.

Rice, B., 1995. Determining the parameters of a rate 1/n convolutional encoder over GF (q). Proceedings of the 3rd International Conference on Finite Fields and Applications, July 11-14, 1995, Glasgow.

Shu, L., S. Lin and D.J. Costello, 2005. Error Control Coding. 2ns Edn., Prentice Hall, New Jersey, USA., pp: 67-95.

Valembois, A., 2001. Detection and recognition of a binary linear code. Discrete Applied Math., 111: 199-218.