Complexity of the Search Algorithm for Some Types of Hash-Coding Schemas

Levon Aslanyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: lasl@sci.am

Hayk Danoyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: hed@ipia.sci.am

ABSTRACT
The exact algorithm of finding sets of "nearest neighbors" in hypercube subsets, using compact blocks and hash functions is known as the Elias algorithm. Perfect codes are the basic structure of composing the balanced block partition of the cube. The problem is investigated not only for perfect codes but also for extensions. This paper integrates the results on hash-coding schemas associated to coverings by intersecting spheres of the same radius, with the base case of perfect codes. These coverings can be obtained via quasi-perfect codes. We consider the mentioned algorithm for the cases of extended Hamming codes and two error-correcting primitive BCH codes of length $2^m - 1$ for odd m as. A formula of time complexity of the algorithm is obtained for these cases.

Keywords
Best match, Nearest neighbor, quasi-perfect codes

1. INTRODUCTION
Let $E = \{0,1\}$. Consider Cartesian degree E^n, which is known as the set of vertices of n-dimensional unit cube. For any $x, y \in E^n$ denote by $d(x,y)$ the Hamming distance between vectors x and y. For an arbitrary $x \in E^n$ denote by $S_r^n(x)$ the sphere of radius r, centred at x i.e., $S_r^n(x) = \{y / y \in E^n, d(x,y) \le r\}$ and by $O_r^n(x)$ denote the shell of radius r, centred at x i.e., $O_r^n(x) = \{y / y \in E^n, d(x,y) = r\}$. We will denote by $car(x)$ the carrier of vector $x = (x_1, \dots, x_n)$ then $car(x) = \{i / x_i = 1, i = 1, \dots, n\}$. Denote by $w(x)$ the weight of vector x i.e., $w(x) = \sum_{i=1}^n x_i$.

Let us have a subset $F \subseteq E^n$ and a vector $x \in E^n$. Let us consider the algorithmic problem of finding the set of all "nearest neighbors" of F to x. More precisely it is required to find the set $F_x = \{y \in F / d(x,y) = d(x,F)\}$. To propose an algorithm for solving the problem of nearest neighbors in the application level, hash coding schemes are considered [1-2]. A brief description of such schemes is brought below.

Hash function is defined as a function $h: E^n \to V$ where $V = \{v_1, \dots, v_N\}$ is a finite set of N elements [1]. Usually cases are considered when $V = E^k$, $k \le n$. The subset F is represented as a union of N disjoint sets (lists). Denote by B_i the set $\{x \in E^n / h(x) = v_i\}$. The i-th list L_i stores those vectors belonging to F, which have the same hash value, i.e., $L_i = \{x \in F / h(x) = v_i\}$ or $L_i = B_i \cap F$, $i = 1, \dots, N$. Hash coding scheme is called balanced if $|B_i| = 2^n / N$.

The Elias algorithm [2] considers blocks B_i ordering them by their distances at vector x. Mention that we must have an efficient method to find all blocks $B_{j_1}, B_{j_2}, \dots, B_{j_{s(j)}}$ located at distance j from x if such blocks exist. After the step of ordering, the algorithm examines the lists $L_{j_1}, L_{j_2}, \dots, L_{j_{s(j)}}$ one after the other by increase of j_t. Let of the best match distance be denoted by δ (also the current value of the best match distance in the algorithm). Due to $F \neq \emptyset$ initialization of δ will happen in some step. Now, if the current values obey $\delta < j$, the algorithm stopes the work. All blocks with higher distances than δ at x do not need to be examined. In the reminder case $\delta \ge j$, examining nonempty list L_{j_t} the algorithm can change the best match distance δ, also refreshing the current best match set, or the δ will remain unchanged and the current best match set will be updated. The pseudocode of the algorithm is brought below:

```
Elias Algorithm      comment: n is the word length,
                              N is the number of blocks
input x, F,          comment: F ≠ ∅
integer δ = ∞,       comment: current best match distance
set S = ∅,           comment: S-is the current set of vectors
                              of F located at distance δ from x
integer j = −1,      comment: current distance from x of
                              blocks under consideration
while(j < δ)
    {j++,
     if(s(j) ≠ 0)    comment: s(j) is the number of blocks
                              located at distance j from x
      for(integer i=0, i<s(j), i++)
        {if(L_{j_i} ≠ ∅)comment: start examine the list
                              L_{j_i}, i-th list with j distanse block
           if(δ ≤ d(x, L_{j_i}))  comment: δ is unchanged
           S = S ∪ (O_δ^n(x) ∩ L_{j_i})  comment: O_δ^n(x) is the
                                          δ neighborhood of x
         else
         {S = O_δ^n(x) ∩ L_{j_i},   comment: δ is changed
              δ = d(x, L_{j_i})
          }
        }
    }
return S,   comment: S = F_x,  δ = d(x, F)
```

By the complexity of the algorithm we mean the average number of examined lists over all files and queries, supposing that:

a) each vector $x \in E^n$ equally likely can be requested,

b) each vector $z \in E^n$ independently appears in F with the same probability p. This gives probabilistic distribution over the set of subsets of E^n.

It is proposed [2,3], that the algorithm is optimal (in above defined sense), when the blocks are isoperimetric sets, the particular case of which is the sphere. Therefore, we consider the coverings of unit cube by nonintersecting spheres of the same radii. As such coverings exist in two simple cases of parameters [4-6], the algorithm and related hash coding schemas are extended to

I) coverings by intersecting spheres of the same radii,

II) coverings by non-intersecting spheres of different radii.

2. PRELIMINARIES

2.1. Definitions and Results from Coding Theory

A nonempty subset C of E^n we call a code [4,5] (in general for codes some prescribed properties take place). The code C is linear if C is a linear subspase of E^n. We will consider only binary codes. Due to the binary nature, the considered C is linear when: $\forall c_1, c_2 \in C \Longrightarrow c_1 + c_2 \in C$, where mod2 summation is applied. Denote by d_C the minimum distance of code C i.e., $d_C = \min\limits_{\substack{c_1,c_2 \in C \\ c_1 \neq c_2}} d(c_1, c_2)$.The packing radius [4,5] of C is called the following nonnegative integer: $r_C = [(d_C - 1)/2]$. Denote by R_C the covering radius of the code C, i.e., $R_C = \max\limits_{x \in E^n} \min\limits_{c \in C} d(x, c)$. In the sequel, not to make a confusion, we use notations d, r and R instead of d_C, r_C and R_C respectively. We say that we have a code $C[n, k, d]R$ if the code C is linear, have dimension k, code length n, minimum distance d and covering radius R. When the code is nonlinear (or it is not known whether the code is linear or not) we use the notation $C(n, M, d)R$ instead, where $M = |C|$. We also use this for linear codes as the second alternative notation.

For $x \in E^n$ the coset of linear code C is called the set $x + C = \{x + c / c \in C\}$. As it is known [4] two different cosets do not intersect, and their union covers the space E^n. We denote by G_C the generator matrix of the linear code $C[n, k]$, with rows forming a basis of code C. Let us denote by H_C the parity check matrix of linear code C. Recall that H_C is $(n - k) \times k$ matrix and for H_C holds the relation $c \in C \Leftrightarrow H_C c^T = 0$. When appropriate we will use notations H and G instead of H_C and G_C respectively. For $x \in E^n$ denote by $A_i(x)$ the number of codewords of C located at distance i from x. The nonnegative integers $A_0^C, A_1^C, \dots, A_n^C$, where $A_i^C = |\{c \in C / w(c) = i\}|$ are called weight spectra of code C. Let us denote by $W_C(x, y)$ the weight enumerator of code C: $W_C(x, y) = \sum_{i=0}^{n} A_i^C x^{n-i} y^i$. Consider weight enumerators depending only on one variable i.e., $W_C(x) = \sum_{i=0}^{n} A_i^C x^i$.

Recall that the dual code $C^{\perp}$ of $[n, k]$ code C is defined as $C^{\perp} = \{y / \langle c, y \rangle = 0, \forall c \in C\}$. It is known [4] that $C^{\perp}$ is $[n, n - k]$ code, and $H_C = G_{C^{\perp}}$. We need the Mac-Williams theorem [4]:

Theorem 1. (Mac-Williams). For linear code C and for its dual code $C^{\perp}$ the following equality takes place:

$$W_C(x, y) = \frac{1}{|C^{\perp}|} W_{C^{\perp}}(x + y, x - y). \quad (1)$$

Substituting x = 1 in (1) and applying the binomial's theorem we get.

$$\sum_{i=0}^{n} A_i^C y^i = \frac{1}{|C^{\perp}|} \sum_{i=0}^{n} A_i^{C^{\perp}} \left(\sum_{j=0}^{n} \left(\sum_{l=0}^{j} (-1)^j \binom{n-i}{j-l} \binom{i}{l} \right) y^j \right).$$

Equalizing the corresponding coefficients, we get:

$$A_j^C = \frac{1}{|C^{\perp}|} \sum_{i=0}^{n} A_i^{C^{\perp}} K_j^n(i),$$

where $K_j^n(i)$ denotes the Kravchouk polynomial of degree j [4,5] i.e. $K_j^n(i) = \sum_{l=0}^{j} (-1)^j \binom{n-i}{j-l} \binom{i}{l}$.

We also need the following theorem, which is proved in [4]:

Theorem 2. Let C be an $[n, k]$ code.Then the arbitrary r columns of code matrix X_C (a code matrix is a matrix, the rows of which are codewords of a code) of code C, where $0 \leq r \leq d_{C^{\perp}} - 1$ contain each vector of lenght r exactly 2^{k-r} times.

2.2 Some Types of Coverings

The code C is called perfect [4,5], if $r_C = R_C$. It is known [4,5] that in binary space nontrivial perfect codes can have only the following two parameter sets:

(I) $(2^m - 1, 2^{2^m - m - 1}, 3)1$,

(II) $(23, 2^{11}, 7)3$.

Here (I) corresponds to the parameters of linear Hemming codes and (II) refers to the case of Golay codes, that are also linear.

Let us consider some generalizations of perfect codes. Let we have a code C, with minimum distance d represented as $2t + 1$ or $2t + 2$ (for odd and even d correspondigly). And we suppose that the covering radius $R \leq t + 1$. Let us denote $D = \{x / d(x, C) \geq t\}$. For $x \in E^n$ denote by $A_i(x)$ the number of codewords of C located at distance i from x. For $x \in D$ denote $a(x) = A_t(x) + A_{t+1}(x)$. Note that $A_t(x) = 1$ or 0. Having $d_C \geq 2t + 1$ and $R_C \leq t + 1$, we may reduce that $a(x) \leq \left[\frac{n+1}{t+1}\right]$. Denote by a the average value of $a(x)$ for all $x \in D$. Then $a = \frac{\sum_{c \in C} |O_t^n(c) \cup O_{t+1}^n(c)|}{2^n - |C| \sum_{i=0}^{t-1} \binom{n}{i}} = \frac{|C| \left(\binom{n}{t} + \binom{n}{t+1} \right)}{2^n - |C| \sum_{i=0}^{t-1} \binom{n}{i}}$.

The code C will be called nearly perfect [4,5] if a(x) achieves the possible maximaum value $\left[\frac{n+1}{t+1}\right]$ for all $x \in D$, i.e., for nearly perfect codes the following equality takes place:

$$|C| \left(\sum_{i=0}^{t-1} \binom{n}{i} + \frac{\binom{n}{t} + \binom{n}{t+1}}{\left[\frac{n+1}{t+1}\right]} \right) = 2^n.$$

The following parameter sets of nearly perfect codes are known:

(III) $(2^m - 2, 2^{2^m - m - 2}, 3)2$,

(IV) $(2^{2m} - 1, 2^{2^{2m} - 4m}, 5)3$.

Here (III) corresponds to the parameters of shortened Hemming codes and (IV) corresponds to parameters of punctured Preparata codes. In [7] it is proved that nearly perfect codes can have only one mentioned parameter sets. The code C will be called strongly uniformly packed if $a(x) = a$ for all $x \in D$ [6]. The parameters of strongly uniformly packed codes are known too [5].

The code C will be called quasi-perfect if $R = r + 1$ [4,5]. Many families of quasi perfect codes are known for the covering radius ≤ 4 [8-13] but the general problem of existence of quasi-perfect codes by the given parameters isn't completely solved yet [8]. Also the nearly perfect codes appear as a special class of quasi-perfect codes.

Let $i \geq 1$ and $R_1, \dots, R_i$ be integers, $C = \bigcup_{j=1}^{i} C_i$. Code C will be called perfect i radius code if the spheres with radii $R_1, \dots, R_i$ respectively centered at points of code sets $C_1, \dots, C_i$ do not intersect and their union covers the whole space [5]. These structures are another candidate that we may apply in model of best match search below, but there are not known exhausting results also about the existence of such codes [5].

When the geometrical interpretation of spherical covers is considered in the models of search of similarities, besides the perfect codes their other possible extensions can be considered and applied, such as nearly perfect codes, strongly uniformly packed codes, quasi perfect codes or coverings by spheres with different radii [5], etc. (2)

2.3 Coset Weight Distribution of Extended Hamming Code

Let us consider the extended Hamming code, which we denote by $\widehat{\mathcal{H}_m}$. It is known [4], that $\widehat{\mathcal{H}_m}$ is $[2^m, 2^m - m - 1, 4]2$ quasi-perfect code, and its parity check matrix is:

$$H_{\widehat{\mathcal{H}_m}} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 1 \end{pmatrix} \quad (3)$$

This is also the generator matrix for the code $\widehat{\mathcal{H}_m^{\perp}}$.

Using (3) we can calculate the weight spectra of $\widehat{\mathcal{H}_m^{\perp}}$, which are brought in Table 1

i	$A_i^{\mathcal{H}_m^{\perp}}$
0	1
2^{m-1}	$2^{m+1}-2$
2^m	1

Table 1

Applying the Mac-Williams's theorem to codes $\widehat{\mathcal{H}_m}$ and $\widehat{\mathcal{H}_m^{\perp}}$ we get:

$$A_j^{\widehat{\mathcal{H}_m^{\perp}}} = \frac{1}{2^{m+1}}\left(K_j^n(0) + (2^{m+1}-2)K_j^n(2^{m-1}) + K_j^n(2^m)\right). \quad (4)$$

It could be obtained [4] that the code has three types of cosets:

(a) $\widehat{\mathcal{H}_m}$,

(b) $e_i + \widehat{\mathcal{H}_m}$, where $car(e_i) = \{i\}$, $i \in \{1, \dots, n\}$,

(c) $g_i + \widehat{\mathcal{H}_m}$, where $car(g_i) = \{1, i\}$, $i \in \{2, \dots, n\}$.

Let us consider the case b. Denote by $L_{e_i} = \widehat{\mathcal{H}_m} \cup \left(e_i + \widehat{\mathcal{H}_m}\right)$. Applying the Mac-Williams's theorem, we get

$$W_{e_i+\widehat{\mathcal{H}_m}}(x,y) = \frac{1}{\left|L_{e_i}^{\perp}\right|} W_{L_{e_i}^{\perp}}(x+y, x-y) - \frac{1}{\left|\widehat{\mathcal{H}_m^{\perp}}\right|} W_{\widehat{\mathcal{H}_m^{\perp}}}(x+y, x-y) \quad (5)$$

Code-words of the code $L_{e_i}^{\perp}$ are those codewords of $\mathcal{H}_{\mathrm{m}}^{\perp}$, for which $\langle \mathrm{x}, \mathrm{e_i} \rangle = 0$. As $w(e_i) = 1$ then it follows from theorem 2 that the number of such codewords is 2^m. So weight spectra of the code $L_{e_i}^{\perp}$ are brought in Table 2.

From Table 2 we get that:

$W_{L_{e_i}^{\perp}}(x,y) = x^{2^m} + (2^m - 1)x^{2^{m-1}}y^{2^{m-1}}$.

Weight spectra of the code $\mathrm{L_{e_i}^{\perp}}$

i	$A_i^{L_{e_i}^{\perp}}$
0	1
2^{m-1}	$2^m - 1$

Table 2

Replacing the latter in (5), and keeping in mind that $\left|L_{e_i}^{\perp}\right| = 2^m$, $\left|\widehat{\mathcal{H}_m^{\perp}}\right| = 2^{m+1}$ we get.

$$W_{e_i+\mathcal{H}_m}(x,y) = \frac{1}{2^{m+1}}\left((x+y)^{2^m} - (x-y)^{2^m}\right) \quad (6)$$

Replacing $\mathrm{x} = 1$ in (6), and applying the binomial's theorem,

$$A_j^{e_i+\widehat{\mathcal{H}_m}} = \frac{1}{2^{m+1}}\binom{2^m}{j}\left(1 - (-1)^j\right) \quad (7)$$

Similarly, for the case c we get:

$$A_j^{g_i+\widehat{\mathcal{H}_m}} = \quad (8)$$

$$= \frac{1}{2^{m+1}}\left(\binom{2^m}{j}\left(1 + (-1)^j\right) - 2K_j^{2^m}(2^{m-1})\right).$$

2.4 Coset Weight Distribution of Uniformly Packed Codes

Later we need the coset weight distributions of two error correcting BCH codes for length $n = 2^{2s+1} - 1$, and for the Golay code. As these codes can be considered as uniformly packed codes [8], then we can find the mentioned distributions by the method, which is brought in [13]:

Theorem 3. Let C be a uniformly packed code with parameters $a_1, \dots, a_R$. Then the polynomial $L_C(x) = \sum_{i=0}^{R} a_i K_i^n(x)$ has R distinct integer roots between 0 and n.

Let us denote those roots by $x_1, \dots, x_R$. Mention that if C is a uniformly packed code containing a zero vector then there exists a uniformly packed code with the same parameters and minimum weight β, where $0 \le \beta \le R$, which we will denote by C_β.

From Theorem 3 and its proof it follows:

Theorem 4. [13]. For the weight function of the uniformly packed code C_β the following equality takes place

$$W_{C_\beta}(x) = \frac{(1+x)^n}{L(0)} + \sum_{i=1}^{R} B_{x_i}^{\beta}(1+x)^{n-x_i}(1-x)^{x_i}. \quad (9)$$

In (9) $B_{x_i}^{\beta}$-s are constants, which we can calculate from (1) by equalizing the corresponding coefficients in the left and right sides and assuming that we know the first R coefficients of $W_{C_\beta}(x)$.

In other words, to find the coefficients $B_{x_i}^{\beta}$, we must solve the corresponding linear system of R equations with R variables.

From Theorem 4 it follows that:

$$A_j^{C_\beta} = \frac{\binom{n}{j}}{\sum_{i=0}^{R} a_i \binom{n}{i}} + \sum_{i=1}^{R}\left(B_{x_i}^{\beta} K_j^n(x_i)\right). \quad (10)$$

Consequently to know the coset weight distribution of uniformly packed code, we must calculate only the first R coset weights, which are $A_0^{C_\beta}, A_1^{C_\beta}, \dots, A_{R-1}^{C_\beta}$.

2.5 Two Error-correcting BCH Codes

Let us denote a finite field of q elements (where q is a power of a prime number) by F_q. We will consider finite fields with characteristic 2. Denote by α the primitive element of the field F_q.Consider the set of formal polynomials $F_q[x]$ with coefficients from the field F_q. As it is known [4], the factor ring $R[x] = F_q[x]/(x^n - 1)$ is a ring of principal ideals, i.e., each ideal in $R[x]$ is principal. An $[n,k]$ code C will be called cyclic code if C is linear, and if from $c = (c_1, c_2, \dots, c_n) \in C$ it follows that $(c_n, c_1, \dots, c_{n-1}) \in C$. We can correspond to each vector $(c_1, c_2, \dots, c_n)$ the polynomial $c_1 + c_2 x + \cdots + c_n x^{n-1}$, so we can consider a code as the subset of $R[x]$. It is known [4], that each cyclic code is an ideal of $R[x]$, i.e., there is a unique polynomial $g(x)$ such that $\forall c(x) \in C\ \exists f(x)\ c(x) = f(x)g(x)$, where multiplication is taken in $R[x]$.

Two error correcting BCH codes are defined as cyclic codes for lengths $n = 2^m - 1$ [4,5] with generator polynomial:

$g(x) = LCM\{M_\alpha(x), M_{\alpha^3}(x)\}$,

where by $M_{\alpha^i}(x)$ is denoted the minimal polynomial of the element α^i. These codes have dimension $2^m - 2m - 1$ and minimum distance equal to5 [4]. It is known that two error correcting BCH codes are quasi-perfect codes [4,13]. The weight distribution of BCH codes was calculated in [4,13].For odd m two error correcting BCH codes are also uniformly packed [8] with parameters $a_0 = a_1 = 1$, $a_2 = a_3 = \frac{6}{n-1}$. Roots of $L_{\mathfrak{B}}(x)$ are $x_1 = \frac{n+1}{2} - \sqrt{\frac{n+1}{2}}$, $x_2 = \frac{n+1}{2}$ and $x_3 = \frac{n+1}{2} + \sqrt{\frac{n+1}{2}}$. It is known, that there are four distinct weight distributions.

For even m 2 error correcting BCH codes are not uniformly packed. It is proved that there are eight distinct coset weight distributions in this case, which are brought in [13].

3. COMPLEXITY OF THE ALGORITHM

Suppose we have an $[n,k]$ code C with covering radius R and $C = \{c_1, c_2, \dots, c_{2^k}\}$. We define a hash function $h: E^n \longrightarrow C$, associated to the code C in the following way:

$$h_C(x) = \{c_i / \ d(x,c_i) = \min_{c \in C}\{d(x,c)\}\}. \quad (11)$$

As it follows from (11), $h_C(x)$ could be a multivalued function because the blocks B_i are spheres of radius R, and they can intersect (recall that $B_i = \{x \in E^n / h_C(x) = c_i\}$, $i \in$

$\{1, \dots, 2^k\}$). When the code C is perfect the mentioned blocks do not intersect and their union covers the unit cube. The formula below for complexity of algorithm is brought for the case corresponding to Hamming code. We also consider hash functions associated to codes in some sense "near" to perfect codes. Such property has also the so called quasi-perfect codes. Indeed, the algorithm is proposed for balanced hash coding schemes, where different blocks B_i do not intersect, but we may also consider the algorithm for the case of intersecting blocks. In this case, when blocks intersect we create the list in a similar way to the basic case and then these lists are also intersecting. Repeated elements bring some redundancy (in terms of memory). To write a formula of complexity of the algorithm, for $x \in E^n$ let us consider the following table:

x	p_1 F_1	p_2 F_2	...	$p_{2^{2^n}}$ $F_{2^{2^n}}$	probability subset
B_1	a_{11}^x	a_{12}^x	...	$a_{12^{2^n}}^x$	
B_2	a_{21}^x	a_{22}^x	...	$a_{22^{2^n}}^x$	
⋮	⋮	⋮	...	⋮	
B_{2^k}	$a_{2^k1}^x$	$a_{2^k2}^x$	...	$a_{2^k2^{2^n}}^x$	

Blocks

Table 3

$F_1, F_2, \dots, F_{2^{2^n}}$ are all subsets of vertices of unit cube and each F_i could be generated with the corresponding probability p_i. We will use the values a_{ij}^x putting them in the cells corresponding to block B_i and subset F_j, where

$$a_{ij}^x = \begin{cases} 1 & if\ B_i is\ considered\ in\ case\ of\ set\ F_i\ and\ vertex\ x, \\ 0 & otherwise. \end{cases}$$

As we mentioned in Section 1, the complexity of algorithm will be represented as

$\alpha(h_C) = \frac{1}{2^n} \sum_{x \in E^n} \sum_{1 \le i \le 2^k} \sum_{1 \le j \le 2^{2^n}} p_j a_{ij}^x$.

Let us denote $\Phi_x(B_i) = \sum_{1 \le j \le 2^{2^n}} p_j a_{ij}^x$. As we can see $\Phi_x(B_i)$ is the probability that the block B_i will be considered by the algorithm when the vector x is requested. Then

$\alpha(h_C) = \frac{1}{2^n} [\sum_{x \in E^n} \sum_{1 \le i \le 2^k} \Phi_x(B_i)]$,

It is easy to understand that for a fixed query x the block B_i will be examined if the sphere $S^n_{d(x,B_i)-1}$ does not contain any vector belonging to F. In that case all blocks B_l such that $d(x, B_l) \le d(x, B_i) - 1$, will be examined. Let j vary over all possible distances between vector x and blocks B_i. Denote by $T_x(j)$ the number of blocks located at distance $\le j$ from vector x, then

$$\alpha(h_C) = \frac{1}{2^n} \sum_{x \in E^n} \sum_{0 \le j \le n} T_x(j) V(j). \quad (12)$$

where $V(j)$ denotes the probability that the nearest vector in F is located at distance j from x. Recall that [2]

$V(j) = (1 - (1-p)^{\binom{n}{j}})(1-p)^{\sum_{l=0}^{j-1}\binom{n}{l}}$.

As $d(x, C_i) = w(x + c_i)$, then the number of vectors located at distance i is equal to A_i^{x+C}. The sphere with centre c_i and radius R will be located in a distance $\le$ j from vector x if and only if $d(x, c_i) \le j + R$. Therefore

$$T_x(j) = \sum_{i=0}^{j+R} A_i^{x+C}. \quad (13)$$

We consider that $A_i^{x+C} = 0$ when $i > n$.

Now let us consider the $[2^m, 2^m - m - 1, 4]2$ extended Hamming code $\widehat{\mathcal{H}_m}$. Recall that the extended Hamming code has three types of cosets enumerated in Section 2.3. Keeping in mind this and the fact that each coset contains 2^{2^m-m-1} vectors and the number of cosets of the first, second and third types is respectively equal to 1, 2^m and $2^m - 1$ we get:

Theorem 5. For the code $\mathfrak{B}_m$ for odd m the complexity of the algorithm is: $\alpha(h_{\widehat{\mathcal{H}_m}}) =$

$$\sum_{0 \le j \le 2^m} V(j) \left(\sum_{i=0}^{j+2} \left(\frac{1}{2^{m+1}} A_i^{\widehat{\mathcal{H}_m}} + \frac{1}{2} A_i^{e_i + \widehat{\mathcal{H}_m}} + \frac{2^m - 1}{2^{m+1}} A_i^{g_i + \widehat{\mathcal{H}_m}} \right) \right).$$

As we mentioned for odd m $\mathfrak{B}_m$ has four types of coset. Keeping in mind this and calculating the number of each type, we will get the following:

Theorem 6. For the code $\mathfrak{B}_m$ for odd m the complexity of the algorithm is:

$$\alpha(h_{\mathfrak{B}_m}) = \sum_{0 \le j \le 2^m - 1} V(j) \sum_{i=0}^{j+3} \left(\frac{1}{2^{2m}} A_i^0 + \frac{2^m - 1}{2^{2m}} A_i^1 + \right.$$
$$\left. + \frac{(2^m - 1)(2^{m-1} - 1)}{2^{2m}} A_i^2 + \frac{2^{2m-1} + 2^{m-1} - 1}{2^{2m}} A_i^3 \right), \quad (15)$$

where by $A_i^j \quad j = 0, .., 3$ is denoted the number of codewords of weight i of coset of minimum weight j.

4. NUMERIC RESULTS

Even having formulas, it is hard to imagine the practical complexities of the things. Two error correcting BCH codes of length $n = 2^{2s+1}$ and Golay code show the tendencies of complexity by n and p. Numerical values of the parameters are easily computable which gives the general complexity picture - average number of considered blocks, and percentage relation of considered elements and cardinality of subset (the tables of numerical results we do not bring due to shortage of place in the paper).

REFERENCES

[1] D. E. Knuth, The Art of Computer Programming, vol. 3, Sorting and Searching, second edition, Eddison-Wesley, 780p., 1998.

[2] R. L. Rivest. On The Optimality of Elias's Algorithm for Performing Best-Match Searches, *Information Processing 74*, North-Holland Publishing Company, 678-681, 1974.

[3] L. H. Aslanyan, The discrete isoperimetry problem and related extremal problems for discrete spaces, *Problemi Kibernetiki*, 36, pp. 85-127, 1979.

[4] F. J. Mac-Williams, N. J. Sloane, The theory of error-correcting codes, North Holland Publishing Company, 762p., 1977.

[5] G. Cohen, I. Honkala, S. Litsyn, A. Lobstein, Covering Codes, *North-Holland Mathematical Library*, vol. 54, 542p. 1997.

[6] V. A. Zinov'ev and V. K. Leont'ev, The nonexistence of perfect codes over Galois fields, *Problems of Control and Info. Theory*, 2(2), pp. 123-132, 1973.

[7] K. Lindstrom, The Nonexistence of Unknown Nearly Perfect Binary Codes. *Ann. Univ. Turku.*, Ser. A, No.169, pp. 3-28 1975.

[8] T. Baicheva, I. Bouyukliev, S. Dodunekov, Binary and ternary linear quasi-perfect codes with small dimensions, *IEEE Transactions of Information Theory*, vol. 54, No 9, pp. 4335-4339, 2008

[9] E. M. Gabidulin, A. A. Davydov, and L. M. Tombak, Linear codes with covering radius 2 and other new covering codes, *IEEE Transactions of Information Theory*, vol. 37, no. 1, pp. 219–224, Jan. 1991.

[10] A. A. Davydov and L. M. Tombak, Quasiperfect linear binary codes with distance 4 and complete caps in projective geometry, *Probl. Inf. Transm.*, vol. 25, no. 4, pp. 265–275, 1989.

[11] A. A. Davydov and A. Yu. Drozhzhina-Labinskaya, Constructions, families, and tables of binary linear covering codes, *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1270–1279, Jul. 1994.

[12] T. Etzion, B. Mounits, Quasi-Perfect Codes With Small Distance, *IEEE Transactions of Information Theory*, Vol. 51, No. 11, pp. 3938-3946, November 2005.

[13] T. Etzion and G. Greenberg, Constructions for perfect mixed codes and other covering codes, *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 209–214, Jan. 1993.