On the Information Function of an Error-Correcting Code

Tor Helleseth, *Senior Member, IEEE*, Torleiv Kløve, *Senior Member, IEEE*, and Vladimir I. Levenshtein, *Member, IEEE*.

***Abstract*—The information function e_h of a code is the average amount of information contained in h positions of the codewords. Upper and lower bounds on the information function of binary linear codes are given. The average value and variance of the information function over all $[n,k]$ codes are determined.**

***Index Terms*— Information function, linear code, support weight, weight hierarchy.**

I. Introduction

We introduce and study the information function e_h of a code. It is defined as the average amount of information contained in h positions of the codewords. We see possible applications of the function in the study of decoding as well as in cryptographic applications of error-correcting codes.

A possible cryptographic application is the following. A code C to be used for transmission of data is chosen at random from the set of all codes of length n (or some suitable subset, e.g., the set of all permutations of some fixed code). Suppose an intruder is able to observe h of the n positions of a codeword. The expected amount of information he will obtain is e_h. If we, as code designers, want him to get as little information as possible, we must choose codes with as small e_h as possible.

For information set decoding (see e.g., [2, pp. 102–131]) we want to have codes with many information sets, that is, $[n,k]$ codes with many sets of k positions containing all the information in a codeword. Therefore, to some extent, we have a design criterion which is the opposite of the criterion for the cryptographic application.

In this paper we consider upper and lower bounds on the information on binary linear codes.

II. Notations and Background Information

Let $\mathcal{S}^n$ denote the subsets of $\{1,2,\cdots,n\}$, and $\mathcal{S}_h = \mathcal{S}_h^n$ the subsets of $\{1,2,\cdots,n\}$ of size h. For a vector $\boldsymbol{c} = (c_1,c_2,\cdots,c_n)$ and a set $X = \{i_1,i_2,\cdots,i_h\} \in \mathcal{S}_h$, where $1 \le i_1 < i_2 < \cdots < i_h \le n$, we let

$$\boldsymbol{c}_X = (c_{i_1}, c_{i_2}, \ldots, c_{i_h}).$$

For a binary code C of length n and a set $X \in \mathcal{S}^n$, we define the code

$$C_X = \{\boldsymbol{c}_X \mid \boldsymbol{c} \in C\}.$$

For $\boldsymbol{y} \in C_X$, let

$$N_X(\boldsymbol{y}) = \left|\{\boldsymbol{c} \in C \mid \boldsymbol{c}_X = \boldsymbol{y}\}\right|.$$

It is clear that

$$\sum_{\boldsymbol{y} \in C_X} N_X(\boldsymbol{y}) = |C|$$

for any $X \in \mathcal{S}^n$.

If all codewords are equally probable and the intruder observes a vector $\boldsymbol{y}$ in the positions in X, then he obtains

$$\log \frac{|C|}{N_X(\boldsymbol{y})}$$

bits of information. The expected (average) information in h positions is therefore

$$e_h = e_h(C) = \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} \frac{1}{|C_X|} \sum_{\boldsymbol{y} \in C_X} \log \frac{|C|}{N_X(\boldsymbol{y})}. \quad (1)$$

We call e_h the *information function* of the code and $e_0, e_1, \cdots, e_n$ the *information profile*.

We note that by (1) and the convexity of the logarithm we get

$$e_h \ge \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} \log |C_X| \quad (2)$$

with equality if and only if all $N_X(\boldsymbol{y})$ are equal for $\boldsymbol{y} \in C_X$. In the case of a linear $[n,k]$ code C, for any X the code C_X is a linear code of some dimension k_X, and $N_X(\boldsymbol{y}) = 2^{k-k_X}$ for all $\boldsymbol{y} \in C_X$. Hence

$$e_h = \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} k_X.$$

In this paper we study the information function of linear codes.

Let G be a generator matrix for C. Since permuting the positions of C does not change the information profile, we may assume that G has the form $(I_k \mid P)$, where I_k is the $k \times k$ identity matrix. For $X \in \mathcal{S}_h$, G_X denotes the $k \times h$ matrix containing the columns of G in the positions of X. By definition, G_X has rank k_X and its rows generate C_X.

The *support* of a vector $\boldsymbol{c}$ is given by

$$\chi(\boldsymbol{c}) = \{i \mid c_i \neq 0\}.$$

For any code D, $\chi(D)$, the *support of* D, is the set of positions where not all the codewords of D are zero, that is,

$$\chi(D) = \bigcup_{\boldsymbol{c} \in D} \chi(\boldsymbol{c}).$$

The *support weight* of D is $w_S(D) = |\chi(D)|$.

Manuscript received April 22, 1996; revised September 12, 1996. This research was supported by The Norwegian Research Council under Project 107623/420.

T. Helleseth and T. Kløve are with the Department of Informatics, University of Bergen, HIB, N-5020 Bergen, Norway.

V. I. Levenshtein is with the Keldysh Institute for Applied Mathematics, Russian Academy of Sciences, 125047 Moscow, Russia.

Publisher Item Identifier S 0018-9448(97)00787-6.

For an $[n, k]$ code C and any r, where $1 \leq r \leq k$, *the* r*th minimum support weight* is defined by

$$d_r = d_r(C) = \min\big\{w_S(D) \mid D \text{ an } [n, r] \text{ subcode of } C\big\}.$$

In particular, the minimum distance of C is d_1. The *weight hierarchy* of C is the set $\{d_1, d_2, \cdots, d_k\}$. These parameters of a code were first studied by Helleseth, Kløve, and Mykkeltveit [5]. The rth minimum support weight is also known as the r*th generalized Hamming weight* [12]. The weight hierarchy, also known as the length/dimension profile, has been studied in a number of papers the last couple of years; for a bibliography, see [4].

Let

$$C^X = \big\{\boldsymbol{c}_X \mid \boldsymbol{c} \in C \text{ and } \chi(\boldsymbol{c}) \subseteq X\big\}.$$

and let k^X denote the dimension of C^X. Let

$$m_h = m_h(C) = \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} k^X.$$

We note that $k^X \leq k_X$. In particular, $m_h \leq e_h$. The inverse of the weight hierarchy is the *dimension/length profile* $k_1, k_2, \cdots, k_n$ defined by

$$k_h = k_h(C) = \max\big\{k^X \mid X \in \mathcal{S}_h\big\}.$$

It was first studied by Kasami *et al*. [6] and Vardy and Be'ery [11]. We also define

$$e_{hr} = e_{hr}(C) = \big|\big\{X \mid X \in \mathcal{S}_h, k_X = r\big\}\big|$$

and

$$m_{hr} = m_{hr}(C) = \big|\big\{X \mid X \in \mathcal{S}_h, k^X = r\big\}\big|$$

for $r \geq 0$. Then

$$\binom{n}{h} e_h = \sum_{r \geq 0} r e_{hr}$$

and

$$\binom{n}{h} m_h = \sum_{r \geq 0} r m_{hr}.$$

For $X \in \mathcal{S}^n$, let

$$\bar{X} = \{1, 2, \cdots, n\} \backslash X.$$

Simonis [10], using different notation, studied the functions e_{hr} and m_{hr} and gave the following lemma and corollaries.

Lemma 1: Let C be an $[n, k]$ code and $X \in \mathcal{S}^n$. Then

i) $(C^X)^\perp = (C^\perp)_X$.
ii) $\dim C_X + \dim (C^\perp)^X = |X|$.
iii) $\dim C_X + \dim C^{\bar{X}} = k$.

From Lemma 1 one gets a number of corollaries.

Corollary 1: Let C be a linear $[n, k]$ code and $0 \leq h \leq n$. Then

i) $e_{hr}(C) = m_{h,h-r}(C^\perp)$, for $0 \leq r \leq h$.
ii) $e_{hr}(C) = m_{n-h,k-r}(C)$, for $0 \leq r \leq h$.
iii) $e_h(C) + m_h(C^\perp) = h$.
iv) $e_h(C) + m_{n-h}(C) = k$.

Combining i) and ii) in Corollary 1 one gets

Corollary 2: Let C be a linear $[n, k]$ code and $0 \leq r \leq h \leq n$. Then

$$e_{hr}(C^\perp) = e_{n-h,k+r-h}(C)$$
$$m_{hr}(C^\perp) = m_{n-h,k+r-h}(C).$$

Combining iii) and iv) in Corollary 1 one gets

Corollary 3: Let C be a linear code and $0 \leq h \leq n$. Then

$$e_h(C^\perp) = e_{n-h}(C) + h - k$$
$$m_h(C^\perp) = m_{n-h}(C) + h - k.$$

For our further investigation, we may assume, without loss of generality, that the k first positions in an $[n, k]$ code are information positions, i.e., the first k columns in a generating matrix are linearly independent. Hence, C has a generator matrix of the form $(I_k \mid P)$, where P is an $k \times (n - k)$ matrix which we call the *redundancy matrix*.

III. Bounds on e_h

Theorem 1: Let C be an $[n, k]$ code generated by $(I_k \mid P)$ and let $p = \operatorname{rank} P$. If $1 \leq h \leq n$, then

$$e_h \geq \frac{hk}{n} + \frac{hp(n-h)}{n(n-1)}.$$

Proof: We may assume (permuting columns if necessary) that

$$G = (I_k | P_1 | P_2)$$

where P_1 is a $k \times p$ matrix of rank p and P_2 is a $k \times (n-k-p)$ matrix. We choose sets of h columns as follows:

First choose a set X_1 of $i \leq h$ columns from P_1. Let $Y = Y(X_1)$ be a set of $k - i$ columns from I_k such that $X_1 \cup Y$ is a basis for $\mathrm{GF}(2)^k$. Choose a set X_2 of j columns from P_2 and a set X_3 of l columns from Y, where $i + l + j \leq h$. Finally, choose a set X_4 of $h - i - l - j$ columns from I_k not in Y. The set $X = X_1 \cup X_2 \cup X_3 \cup X_4$ has rank at least $i + l$. Since all sets X chosen in this way are distinct, we get

$$\binom{n}{h} e_h \geq \sum_{i=0}^{p} \binom{p}{i} \sum_{j=0}^{n-k-p} \binom{n-k-p}{j} \cdot \sum_{l=0}^{k-i} \binom{k-i}{l} \binom{i}{h-i-j-l} (i+l).$$

We remind the reader about the Vandermonde convolution

$$\sum_{m=0}^{b-a} \binom{b-a}{m} \binom{a}{p-m} = \binom{b}{p}.$$

Using this we get

$$\sum_{i=0}^{p} \binom{p}{i} \sum_{j=0}^{n-k-p} \binom{n-k-p}{j} \sum_{l=0}^{k-i} \binom{k-i}{l} \binom{i}{h-i-j-l} i = \sum_{i=1}^{p} i \binom{p}{i} \sum_{j=0}^{n-k-p} \binom{n-k-p}{j} \binom{k}{h-i-j}$$

$$
\begin{aligned}
&= \sum_{i=1}^{p} p \binom{p-1}{i-1} \binom{n-p}{h-i} \\
&= p \sum_{i=1}^{p} \binom{p-1}{i-1} \binom{(n-1)-(p-1)}{(h-1)-(i-1)} \\
&= p \binom{n-1}{h-1}.
\end{aligned}
$$

Similarly

$$
\sum_{i=0}^{p} \binom{p}{i} \sum_{j=0}^{n-k-p} \binom{n-k-p}{j} \sum_{l=0}^{k-i} \binom{k-i}{l} \binom{i}{h-i-j-l} l = k\binom{n-1}{h-1} - p\binom{n-2}{h-2}.
$$

Combining these we get

$$
e_h \geq \frac{p\binom{n-1}{h-1} + k\binom{n-1}{h-1} - p\binom{n-2}{h-2}}{\binom{n}{h}} = \frac{hk}{n} + \frac{hp(n-h)}{n(n-1)}. \qquad \square
$$

Remark 1: We see from the proof that if C is an $[n, k]$ code generated by

$$
\left(I_k \middle| \begin{matrix} I_p \\ O_{k-p,p} \end{matrix} \middle| O_{k,n-k-p} \right)
$$

where $O_{k-p,p}$ is the $(k-p) \times p$ all-zero matrix and $O_{k,n-k-p}$ is the $k \times (n-k-p)$ all-zero matrix, then

$$
e_h = \frac{hk}{n} + \frac{hp(n-h)}{n(n-1)}
$$

for $1 \leq h \leq n$. In particular, for the $[n, k]$ code generated by the matrix $(I_k \mid O_{k,n-k})$ we get $e_h = hk/n$ for all h and this is the smallest possible value of e_h for an $[n, k]$ code.

Remark 2: If we choose $X_3 = Y$, the resulting X has rank k. Hence, we also get, by the same argument,

$$
e_{hk} \geq \sum_{i=0}^{p} \binom{p}{i} \binom{n-k-p+i}{h-k}.
$$

In particular, for the number of information sets e_{kk} we get $e_{kk} \geq 2^p$.

For an $[n, k, d]$ code with $d \geq 2$, the rank p of a redundancy matrix cannot be too small, and an interesting problem is to find the minimal rank $p(n, k, d)$ of any redundancy matrix of any $[n, k, d]$ code. It is clear that

$$
p(n, k, d) \leq n - k.
$$

On the other hand, by Theorem 1 and Remark 2 above, any lower bound on $p(n, k, d)$ gives rise to a lower bound on e_h and a lower bound on the number of information sets.

Let $K(N, d)$ denote the maximal dimension of a binary linear code of length N and minimum distance d. Similarly, let $T(N, d)$ denote the minimal dimension of a binary linear code of length N and dual distance d. Clearly, if C is an $[N, k, d]$ code, then $C^{\perp}$ is an $[N, N-k]$ code of dual distance d and vice versa. Hence

$$
T(N, d) = N - K(N, d).
$$

Lemma 2: If there exists an $[n, k, d]$ code, then

$$
p(n, k, d) \geq T(k, d).
$$

Proof: Let P be a redundancy matrix for an $[n, k, d]$ code. Any $d-1$ rows of P must be linearly independent, since otherwise a linear combination is zero, and the corresponding linear combination of rows in $(I_k \mid P)$ then gives a codeword of weight less than d. Hence P^T generates a code of length k and dual distance at least d. $\square$

From known lower bounds on $T(N, d)$ due to Singleton, Rao, and Levenshtein (see [9]) we obtain the following bounds on $p(n, k, d)$ for $k \geq d \geq 2$:

1) $p(n, k, d) \geq d - 1$;
2) $p(n, k, d) \geq \log\left(2^{\theta} \sum_{i=0}^{l} \binom{k-\theta}{i}\right)$
   if $d = 2l + 1 + \theta$, where l is an integer and $\theta \in \{0, 1\}$;
3) $p(n, k, d) \geq k - \log L^{(k)}(d)$
   where

$$
L^{(k)}(d) = \begin{cases} L_m^{(k)}(d), & \text{if } d_m(k-1) < d-1 \leq d_{m-1}(k-2) \\ 2L_m^{(k-1)}(d), & \text{if } d_m(k-2) < d-1 \leq d_m(k-1) \end{cases}
$$

where $d_m(k)$ is the smallest root of the Krawtchouk polynomial

$$
K_m^k(z) = \sum_{j=0}^{m} (-1)^j \binom{z}{j} \binom{k-z}{m-j}
$$

of degree m, and

$$
L_m^{(k)}(d) = \sum_{i=0}^{m-1} \binom{k}{i} - \binom{k}{m} \frac{K_{m-1}^{k-1}(d-1)}{K_m^k(d)}.
$$

Now we go on to investigate relations between e_{h+1} and e_h. For $X \in \mathcal{S}_h$ and $Y \in \mathcal{S}_{h+1}$, where $X \subset Y$, let

$$
\kappa_{XY} = k_Y - k_X.
$$

Clearly, $\kappa_{XY} \in \{0, 1\}$. Each $X \in \mathcal{S}_h$ is contained in exactly $n - h$ sets $Y \in \mathcal{S}_{h+1}$, and each $Y \in \mathcal{S}_{h+1}$ contains exactly $h + 1$ sets $X \in \mathcal{S}_h$. Hence

$$
\begin{aligned}
e_{h+1} &= \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} \frac{1}{n-h} \sum_{\substack{Y \in \mathcal{S}_{h+1} \\ X \subset Y}} (k_X + \kappa_{XY}) \\
&= \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} \frac{1}{n-h} \sum_{\substack{Y \in \mathcal{S}_{h+1} \\ X \subset Y}} k_X \\
&\quad + \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} \frac{1}{n-h} \sum_{\substack{Y \in \mathcal{S}_{h+1} \\ X \subset Y}} \kappa_{XY} \\
&= e_h + \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} \frac{1}{n-h} \sum_{\substack{Y \in \mathcal{S}_{h+1} \\ X \subset Y}} \kappa_{XY} \\
&= e_h + \frac{1}{\binom{n}{h}} \sum_{X \in \mathcal{S}_h} \frac{1}{n-h} \\
&\quad \cdot \left| \{Y \mid Y \in \mathcal{S}_{h+1}, X \subset Y, \kappa_{XY} = 1\} \right|.
\end{aligned}
$$

Hence we get the following lemma.

Lemma 3: Let C be an $[n,k]$ code. Then

$$e_{h+1} = e_h + \frac{1}{(n-h)\binom{n}{h}} \cdot \sum_{X \in \mathcal{S}_h} \left|\{Y \mid Y \in \mathcal{S}_{h+1}, X \subset Y, \kappa_{XY} = 1\}\right|.$$

From Lemma 3 we get $e_h \le e_{h-1} + 1$, which can also be seen directly. A more important application of Lemma 3 is to give another good lower bound.

Theorem 2: Let C be an $[n,k]$ code and $1 \le h < n$. Then

$$e_{h+1} \ge e_h + \frac{k - e_h}{n-h}.$$

Proof: Let $X \in \mathcal{S}_h$. The columns of G_X generate a vector space V of dimension k_X. Therefore, there are $k - k_X$ columns in the information positions which are not contained in V. Any of the corresponding positions together with X gives a set $Y \in \mathcal{S}_{h+1}$ containing X and such that $k_Y = k_X + 1$. Hence

$$\left|\{Y \mid Y \in \mathcal{S}_{h+1}, X \subset Y, \kappa_{XY} = 1\}\right| \ge k - k_X.$$

Combining this with Lemma 3 we get

$$\begin{aligned} e_{h+1} &\ge e_h + \frac{1}{(n-h)\binom{n}{h}} \sum_{X \in \mathcal{S}_h} (k - k_X) \\ &= e_h + \frac{k}{(n-h)\binom{n}{h}} |\mathcal{S}_h| - \frac{1}{(n-h)\binom{n}{h}} \sum_{X \in \mathcal{S}_h} k_X \\ &= e_h + \frac{k}{n-h} - \frac{e_h}{n-h}. \end{aligned}$$

□

The bounds in Theorem 2 are best possible in the sense that there exists a code for which $e_{h+1} = e_h + (k - e_h)/(n-h)$ for all h, namely, the code generated by the matrix $(I_k \mid \mathbf{0})$.

Theorem 3: Let C be an $[n,k]$ code. Let $1 \le r \le k$ and $0 \le h \le n$. Then

$$e_h \le k - r + \frac{h}{n} d_r.$$

Proof: Let G be a generator matrix for C with the property that the first r rows of G generate an r-dimensional subcode D_r of C of support weight d_r. Let $X \in \mathcal{S}_h$ and $Y = X \cap \chi(D_r)$. The last $k - r$ rows of G_X have rank at most $k - r$. The first r rows have rank $k_Y(D_r) \le |Y|$. Hence

$$k_X \le k - r + |Y|$$

and we get

$$\begin{aligned} e_h(C) &\le \frac{1}{\binom{n}{h}} \sum_{i=0}^{\min(h,d_r)} \sum_{Y \in \mathcal{S}_i^{d_r}} (k - r + i) \binom{n - d_r}{h - i} \\ &\le \frac{k-r}{\binom{n}{h}} \sum_{i=0}^{\min(h,d_r)} \binom{d_r}{i} \binom{n-d_r}{h-i} \\ &\quad + \frac{1}{\binom{n}{h}} \sum_{i=0}^{\min(h,d_r)} i \binom{d_r}{i} \binom{n-d_r}{h-i} \\ &= k - r + \frac{h}{n} d_r. \end{aligned}$$

□

IV. Some Further Properties of e_{hr}

Some further properties of e_{hr} are given in the next theorems.

Theorem 4: Let C be an $[n,k]$ code and $1 \le h \le n$. Then

i) $e_{hr} = 0$ for $0 \le r < h - k_h(C^\perp)$,
ii) $e_{hr} = 0$ for $r > \min(h,k)$,
iii) $e_{hr} \ge 1$ for $h - k_h(C^\perp) \le r \le \min(h,k)$.

Proof:

i) Let $X \in \{1, 2, \cdots, n\}$. Then

$$h - k_X = \dim (C_X)^\perp = \dim (C^\perp)^X \le k_h(C^\perp)$$

which proves i). Moreover, there exists an X such that

$$\dim (C^\perp)^X = k_h(C^\perp)$$

and so

$$e_{h, h - k_h(C^\perp)} \ge 1.$$

ii) For any $X \in \mathcal{S}_h$, the code C_X is an $[h, k_X]$ code. In particular $k_X \le h$ and $k_X \le k$, and so $e_h \le \min(h,k)$. Moreover, there exists an $Y \in \mathcal{S}_h$ such that $k_Y = \min(h,k)$. This proves ii) and that

$$e_{h, \min(h,k)} \ge 1.$$

iii) There exists a sequence $X = X_1, X_2, \cdots, X_t = Y$ of sets in $\mathcal{S}_h$ such that $|X_i \cap X_{i+1}| = h - 1$ for $i = 1, 2, \cdots, t-1$. Clearly, $|k_{X_i} - k_{X_{i+1}}| \le 1$. Hence, for each r such that $h - k_h(C^\perp) \le r \le \min(h,k)$, there exists an i such that $k_{X_i} = r$. This proves iii). □

For $0 \le r \le k$ and $0 \le i \le n$ let A_i^r denote the number of subspaces of C of dimension r and support weight i. In particular, $1, A_1^1, A_2^1, \cdots, A_n^1$ is the usual weight distribution of C (note that $A_0^1 = 0$).

We next give a lemma and a theorem which both are essentially due to Simonis [10].

Let

$$[r] = \prod_{j=0}^{r-1} (2^r - 2^j)$$

which is the number of ordered bases of an r-dimensional space (over GF (2)), and let

$$\begin{bmatrix} a \\ r \end{bmatrix} = \prod_{j=0}^{r-1} \frac{2^a - 2^j}{2^r - 2^j} = 2^{-r(a-r)} \frac{[a]}{[r][a-r]}$$

which is the number of r-dimensional subspaces of an a-dimensional space.

Lemma 4: Let C be an $[n,k]$ code and $1 \le r \le h \le n$. Then

$$\sum_{s=0}^{r} \begin{bmatrix} k - s \\ k - r \end{bmatrix} e_{hs} = \sum_{i=0}^{n} A_i^{k-r} \binom{n-i}{h}.$$

Lemma 4 gives a set of equations for e_{hr} which can be solved.

Theorem 5: Let C be an $[n,k]$ code and $1 \le r \le h \le n$. Then

$$e_{hr} = \sum_{j=0}^{r} (-1)^j 2^{\binom{j}{2}} \begin{bmatrix} k-r+j \\ j \end{bmatrix} \sum_{i=0}^{n} A_i^{k-r+j} \binom{n-i}{h}.$$

In particular

$$\binom{n}{h} e_h = \sum_{r=1}^{h} r \sum_{j=0}^{r} (-1)^j 2^{\binom{j}{2}} \begin{bmatrix} k-r+j \\ j \end{bmatrix} \cdot \sum_{i=0}^{n} A_i^{k-r+j} \binom{n-i}{h}.$$

Corollary 4: Let C be an $[n,k,d]$ code. If $n-d < h \le n$, then $e_h = k$.

Proof: If $i \ge d$, then $\binom{n-i}{h} = 0$. If $i < d$ and $r - j < k$, then $A_i^{k-r+j} = 0$. Finally, if $i < d$, $r-j = k$, and $A_i^{k-r+j} > 0$, then $j = 0$, $r = k$, $i = 0$. Hence $e_h = k$.

Alternatively, a direct proof goes as follows: Let $X \in \mathcal{S}_h$. Then $|X| = h > n-d$ and so $k - k_X = 0$. □

A similar argument gives the following corollary.

Corollary 5: Let C be an $[n,k,d]$ code. If $n - d_2 < h \le n - d$, then

$$e_{hk} = \binom{n}{h} - \sum_{i=d}^{n-h} A_i \binom{n-i}{h}$$
$$e_{h,k-1} = \sum_{i=d}^{n-h} A_i \binom{n-i}{h}$$
$$e_{hr} = 0, \quad \text{for } r < k-1.$$

In particular

$$e_h = k - \sum_{i=d}^{n-h} A_i \frac{\binom{n-i}{h}}{\binom{n}{h}} < k$$

and

$$e_h \ge k - (2^k - 1) \frac{\binom{n-d}{h}}{\binom{n}{h}}.$$

Combining Theorem 2 and Corollaries 4 and 5, we get the following corollary.

Corollary 6: Let C be an $[n,k,d]$ code. Then

$$0 = e_0 < e_1 < \cdots < e_{n-d} < e_{n-d+1} = e_{n-d+2} = \cdots = e_n = k.$$

Example: Let C be the $[2^k - 1, k]$ simplex code. Then

$$A^r_{2^k - 2^{k-r}} = \begin{bmatrix} k \\ r \end{bmatrix}$$
$$A_i^r = 0, \quad \text{otherwise}.$$

Hence we get

$$e_{hr} = \sum_{j=0}^{r} (-1)^j 2^{\binom{j}{2}} \begin{bmatrix} k-r+j \\ j \end{bmatrix} \begin{bmatrix} k \\ k-r+j \end{bmatrix} \binom{2^{r-j}-1}{h}$$
$$= \sum_{j=0}^{r-1} (-1)^j 2^{\binom{j}{2}} \begin{bmatrix} k \\ r \end{bmatrix} \begin{bmatrix} r \\ j \end{bmatrix} \binom{2^{r-j}-1}{h}.$$

This particular example was also studied by Laksov [8] and Carlitz [1] in a different context. They derived an equivalent expression for e_{hr} (in a different notation).

V. Relations with the Dual Code

Let C be an $[n,k]$ code. As usual, we let $B_0, B_1, \cdots, B_n$ denote the weight distribution of the dual code $C^\perp$. Further, we let $d_1^\perp, d_2^\perp, \cdots, d_{n-k}^\perp$ denote the weight hierarchy of $C^\perp$. In particular, $d_1^\perp$ is the dual distance.

Lemma 5: For all h we have

$$e_{h0} = \binom{B_1}{h}.$$

Proof: This follows immediately from the fact that B_1 is the number of all-zero columns in G. □

In particular, we get

$$e_{10} = B_1$$
$$e_{11} = n - B_1.$$

For $h = 2$ we see that k_X is 1 if the two corresponding columns in G are $\{\mathbf{0}, \boldsymbol{x}\}$ or $\{\boldsymbol{x}, \boldsymbol{x}\}$, where $\boldsymbol{x} \ne \mathbf{0}$. Hence

$$e_{20} = \binom{B_1}{2}$$
$$e_{21} = B_1(n - B_1) + \left(B_2 - \binom{B_1}{2}\right)$$
$$e_{22} = \binom{n}{2} - B_2 - B_1(n - B_1).$$

Lemma 6: For $h < d_1^\perp$ we have $e_h = h$.

Proof: This follows immediately from the fact that any $h < d_1^\perp$ columns in G are linearly independent, that is, $k_X = |X|$ if $|X| < d_1^\perp$. □

Example: For the $[n, n-1, 2]$ even-weight code Lemma 6 gives $e_h = h$ for $h < n$. Further, $e_n = k = n - 1$.

Lemma 7: Let C be an $[n,k]$ code and $d_1^\perp \le h < d_2^\perp$. Then

$$e_h = h - \frac{1}{\binom{n}{h}} \sum_{i=d_1^\perp}^{d_2^\perp - 1} B_i \binom{n-i}{h-i} (h - i + 1).$$

Proof: Let $X \in \mathcal{S}_h$. There is at most one codeword in $C^\perp$ with support in X since two such codewords would generate a subspace of $C^\perp$ of dimension 2 and support weight $\le |X| = h < d_2^\perp$. The support of a codeword in $C^\perp$ of weight $i \le h$ is contained in $\binom{n-i}{h-i}$ sets $X \in \mathcal{S}_h$. For these sets we have

$$k_X = i - 1.$$

For the sets X which contain no codewords we must have $k_X = |X|$. Hence

$$e_h = \frac{1}{\binom{n}{h}} \left\{ \sum_{i=d_1^\perp}^{d_2^\perp - 1} B_i \binom{n-i}{h-i} (i-1) + h \left\{ \binom{n}{h} - \sum_{i=d_1^\perp}^{d_2^\perp - 1} B_i \binom{n-i}{h-i} \right\} \right\}.$$
$$= h - \frac{1}{\binom{n}{h}} \sum_{i=d_1^\perp}^{d_2^\perp - 1} B_i \binom{n-i}{h-i} (h - i + 1). \qquad \square$$

VI. The Minimum Problem as a Programming Problem

Let C be an $[n, k]$ code and G a generator matrix for C. For each $\boldsymbol{v} \in \mathrm{GF}(2)^k$, let $x_{\boldsymbol{v}}$ denote the number of times $\boldsymbol{v}$ appears as a column of G. Further, let V_s be the set of s-dimensional subspaces of $\mathrm{GF}(2)^k$. Helleseth *et al.* [5] introduced a one-to-one correspondence $U \mapsto D_U$ between the spaces in V_s and the $(k-s)$-dimensional subspaces of C such that

$$w_S(D_U) = n - \sum_{\boldsymbol{v} \in U} x_{\boldsymbol{v}}.$$

Using this fact, Theorem 5 can be reformulated as follows: Let $1 \le h \le n$. Then

$$\binom{n}{h} e_h = \sum_{r=1}^{\min(h,k)} r \sum_{j=0}^{r} (-1)^j 2^{\binom{j}{2}} \begin{bmatrix} k-r+j \\ j \end{bmatrix} \cdot \sum_{U \in V_{r-j}} \binom{\sum_{\boldsymbol{v} \in U} x_{\boldsymbol{v}}}{h}. \tag{3}$$

Let

$$\mu(n, k, d, h) = \min \{ e_h(C) \mid C \text{ is an } [n, k, d] \text{ code} \}.$$

Then

$$\mu(n, k, d, h) = \min \frac{1}{\binom{n}{h}} \sum_{r=1}^{\min(h,k)} r \sum_{j=0}^{r} (-1)^j 2^{\binom{j}{2}} \cdot \begin{bmatrix} k-r+j \\ j \end{bmatrix} \sum_{U \in V_{r-j}} \binom{\sum_{\boldsymbol{v} \in U} x_{\boldsymbol{v}}}{h} \tag{4}$$

where the minimum is taken over all $(x_{\boldsymbol{v}})_{\boldsymbol{v} \in \mathrm{GF}(2)^k}$ such that

$$x_{\boldsymbol{v}} \text{ is a nonnegative integer}$$

$$\sum_{\boldsymbol{v} \in \mathrm{GF}(2)^k} x_{\boldsymbol{v}} = n$$

and

$$\sum_{\boldsymbol{v} \in U} x_{\boldsymbol{v}} \le n - d$$

for all $U \in V_{k-1}$. Hence, the determination of $\mu(n, k, d, h)$ is formulated as an integer programming problem.

To determine $\mu(n, k, d, h)$ in general seems to be a difficult problem. We will determine $\mu(n, 2, d, h)$ as an illustration. Let C be an $[n, 2, d]$ code. From (3) we get

$$\binom{n}{h} e_h = 2\binom{n}{h} - \binom{x_{00}+x_{01}}{h} - \binom{x_{00}+x_{10}}{h} - \binom{x_{00}+x_{11}}{h} + \binom{x_{00}}{h}.$$

Without loss of generality, we may assume that

$$x_{01} \ge x_{10} \ge x_{11}.$$

Then

$$x_{00} + x_{01} = n - d \ge x_{00} + x_{10} \ge x_{00} + x_{11}.$$

We obtain $\mu(n, 2, d, h)$, the minimal value of e_h, as follows: Suppose $x_{10} \ge x_{11} + 2$. Then we let $x'_{10} = x_{10} - 1$, $x'_{11} = x_{11} + 1$, $x'_{01} = x_{01} - 1$, $x'_{00} = x_{00} + 1$. Then

$$\begin{aligned}\binom{n}{h}(e_h - e'_h) &= \binom{x_{00}+x_{11}+2}{h} - \binom{x_{00}+x_{11}}{h} - \binom{x_{00}}{h-1} \\ &= \binom{x_{00}+x_{11}+1}{h-1} + \binom{x_{00}+x_{11}}{h-1} - \binom{x_{00}}{h-1} \ge 0.\end{aligned}$$

Similarly, if $x_{01} > x_{10}$ we can let $x'_{01} = x_{01} - 1$, $x'_{00} = x_{00} + 1$, $x'_{10} = x_{10}$, $x'_{11} = x_{11}$, and the value of e_h will not increase. Hence for the minimum we have $x_{01} = x_{10} = x_{11}$ or $x_{01} = x_{10} = x_{11} + 1$, that is
$d = 2t$ (d is even):

$$x_{01} = t, \; x_{10} = t, \; x_{11} = t, \; x_{00} = n - 3t.$$

$d = 2t - 1$ (d is odd):

$$x_{01} = t, \; x_{10} = t, \; x_{11} = t - 1, \; x_{00} = n - 3t + 1.$$

This gives the following results.

Theorem 6:

i) If d is even, then

$$\mu(n, 2, d, h) = 2 - 3\frac{\binom{n-d}{h}}{\binom{n}{h}} + \frac{\binom{n-3d/2}{h}}{\binom{n}{h}}.$$

ii) If d is odd, then

$$\mu(n, 2, d, h) = 2 - 2\frac{\binom{n-d}{h}}{\binom{n}{h}} - \frac{\binom{n-d-1}{h}}{\binom{n}{h}} + \frac{\binom{n-(3d+1)/2}{h}}{\binom{n}{h}}.$$

We have also considered $k = 3$. Numerical results indicate that the minimum is obtained for the following values of the variables (for $d \ge 2$):

d	x_{000}	x_{100}	x_{010}	x_{110}
$4t$	$n-7t$	t	t	t
$4t-1$	$n-7t+1$	t	t	t
$4t-2$	$n-7t+3$	$t-1$	t	t
$4t-3$	$n-7t+4$	$t-1$	t	t

d	x_{001}	x_{101}	x_{011}	x_{111}
$4t$	t	t	t	t
$4t-1$	t	t	t	$t-1$
$4t-2$	t	t	$t-1$	$t-1$
$4t-3$	t	t	$t-1$	$t-2$

This gives the following (conjectured) values for $\binom{n}{h}\mu(n,3,d,h)$:

$d \bmod 4$	$\mu(n,3,d,h)$
0	$3\binom{n}{h} - 7\binom{n-d}{h} + 7\binom{n-3d/2}{h} - 3\binom{n-7d/4}{h}$
-1	$3\binom{n}{h} - 4\binom{n-d}{h} - 3\binom{n-d-1}{h} + 6\binom{n-(3d+1)/2}{h}$ $+\binom{n-(3d+1)/2-1}{h} - 3\binom{n-(7d+3)/4}{h}$
-2	$3\binom{n}{h} - 6\binom{n-d}{h} - \binom{n-d-2}{h} + 4\binom{n-3d/2}{h}$ $+3\binom{n-3d/2-1}{h} - 3\binom{n-(7d+2)/4}{h}$
-3	$3\binom{n}{h} - 4\binom{n-d}{h} - 2\binom{n-d-1}{h} - \binom{n-d-3}{h}$ $+4\binom{n-(3d+1)/2}{h} + 2\binom{n-(3d+1)/2-1}{h}$ $+\binom{n-(3d+1)/2-2}{h} - 3\binom{n-(7d+5)/4}{h}$

In a few cases we can give lower bounds on $\mu(n,k,d,h)$.

Theorem 7: For all n, k, and d we have

$$\mu(n,k,d,1) \geq \frac{d(2^k-1)}{n2^{k-1}}.$$

Proof: The function we want to minimize can be rewritten as

$$\frac{1}{n}\sum_{\boldsymbol{v}\neq \mathbf{0}} x_{\boldsymbol{v}}.$$

By the Plotkin bound

$$\frac{1}{n}\sum_{\boldsymbol{v}\neq \mathbf{0}} x_{\boldsymbol{v}} \geq \frac{d(2^k-1)}{n2^{k-1}}$$

and we have equality if and only $x_{\boldsymbol{v}} = \frac{d}{2^{k-1}}$ for all $\boldsymbol{v} \neq \mathbf{0}$. □

Theorem 8:

i) If $n-d < h \leq n$, then

$$\mu(n,k,d,h) = k.$$

ii) If $n - \lceil 3d/2 \rceil < h \leq n-d$, then

$$\mu(n,k,d,h) \geq k - (2^k-1)\frac{\binom{n-d}{h}}{\binom{n}{h}}.$$

Proof: i) follows directly from Corollary 4. By the Griesmer–Wei bound, $d_2 \geq \lceil 3d/2 \rceil$ and so ii) follows from Corollary 5. □

VII. The Average Information Function

In this section we consider the average value of e_{hr} and e_h over all $k\times n$ generator matrices, or equivalently, over all $[n,k]$ codes. We start with some technical lemmas.

Lemma 8: The number of binary $k\times h$ matrices of rank r is

$$\begin{bmatrix} k \\ r \end{bmatrix}\begin{bmatrix} h \\ r \end{bmatrix}[r].$$

The lemma is well known and can, e.g., be proved as follows: We count the number of linear mappings $\mathrm{GF}(2)^k \to \mathrm{GF}(2)^h$ of rank r. The number of kernels is $\left[\begin{smallmatrix} k \\ r \end{smallmatrix}\right]$ and the number of image spaces is $\left[\begin{smallmatrix} h \\ r \end{smallmatrix}\right]$. Finally, an ordered basis of a space of dimension r can be chosen in $[r]$ ways.

Lemma 9: Let P be a binary $k\times l$ matrix of rank s. Then the number $\gamma(k,m,s,t)$ of $k\times m$ matrices Q such that $(P \mid Q)$ has rank $s+t$ is exactly

$$\gamma(k,m,s,t) = 2^{sm}\begin{bmatrix} m \\ t \end{bmatrix}\begin{bmatrix} k-s \\ t \end{bmatrix}[t].$$

Proof: First we observe that any $P' = AP$ for an invertible matrix A gives the same number. Hence we may assume that

$$P = \begin{pmatrix} I_s \\ O_{k-s,s} \end{pmatrix}$$

and Q is of the form

$$\begin{pmatrix} X \\ Y \end{pmatrix}$$

where I_s is the $s\times s$ identity matrix, $O_{k-s,s}$ is the $(k-s)\times s$ all-zero matrix, X is an arbitrary $s\times m$ matrix, and Y is a $(k-s)\times m$ matrix of rank t. Therefore, X can be chosen in 2^{sm} ways, and, by Lemma 8, Y can be chosen in $\left[\begin{smallmatrix} m \\ t \end{smallmatrix}\right]\left[\begin{smallmatrix} k-s \\ t \end{smallmatrix}\right][t]$ ways. □

Theorem 9: The average value of e_{hr} over all $[n,k]$ codes is

$$E(e_{hr}) = \binom{n}{h}\frac{\begin{bmatrix} h \\ r \end{bmatrix}\begin{bmatrix} n-h \\ k-r \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}}2^{r(n-h-k+r)}.$$

Proof: A binary $k\times n$ matrix of rank k can be chosen in $\gamma(k,n,0,k)$ ways. For any of the $\binom{n}{h}$ choices of h positions, there are $\gamma(k,h,0,r)\gamma(k,n-h,r,k-r)$ of those matrices which have rank r in these h positions and thereby contribute to the average. Hence, the average is

$$\binom{n}{h}\frac{\gamma(k,h,0,r)\gamma(k,n-h,r,k-r)}{\gamma(k,n,0,k)}$$

and so the theorem follows from Lemma 9. □

Theorem 10: The average value of e_h over all $[n,k]$ codes is

$$E(e_h) = \frac{1}{\begin{bmatrix} n \\ k \end{bmatrix}}\sum_{r=\max(0,h+k-n)}^{\min(h,k)} r\begin{bmatrix} h \\ r \end{bmatrix}\begin{bmatrix} n-h \\ k-r \end{bmatrix}2^{r(n-h-k+r)}.$$

We have

$$\sum_{r=\max(0,h+k-n)}^{\min(h,k)} \begin{bmatrix} h \\ r \end{bmatrix}\begin{bmatrix} n-h \\ k-r \end{bmatrix}2^{r(n-h-k+r)} = \begin{bmatrix} n \\ k \end{bmatrix}.$$

Hence we get the following corollaries.

Corollary 7: If $h \leq k$, then the average value of e_h over all $[n,k]$ codes is

$$E(e_h) = h - \frac{1}{\begin{bmatrix} n \\ k \end{bmatrix}}\sum_{r=1}^{\min(h,n-k)} r\begin{bmatrix} h \\ r \end{bmatrix}\begin{bmatrix} n-h \\ k-h+r \end{bmatrix}2^{(h-r)(n-k-r)}.$$

Corollary 8: If $h \geq k$, then the average value of e_h over all $[n, k]$ codes is

$$E(e_h)=k-\frac{1}{\begin{bmatrix} n \\ k \end{bmatrix}}\sum_{r=1}^{\min(k,n-h)} r\begin{bmatrix} h \\ k-r \end{bmatrix}\begin{bmatrix} n-h \\ r \end{bmatrix}2^{(k-r)(n-h-r)}.$$

Examples: If $k \leq n-1$ then

$$E(e_{n-1}) = k - \frac{2^k-1}{2^n-1}. \tag{5}$$

If $k \leq n-2$ then

$$E(e_{n-2}) = k - 3\frac{2^k-1}{2^n-1} + \frac{(2^k-1)(2^{k-1}-1)}{(2^n-1)(2^{n-1}-1)}.$$

Lemma 10: For $0 \leq b \leq a$ we have

$$2^{b(a-b)} \leq \begin{bmatrix} a \\ b \end{bmatrix} \leq c^{-1}2^{b(a-b)}$$

where

$$c = \prod_{i=1}^{\infty}\left(1-\frac{1}{2^i}\right) \approx 0.2887.$$

Proof: The lower bound follows from the fact that

$$\frac{2^{a-i}-1}{2^{b-i}-1} \geq \frac{2^{a-i}}{2^{b-i}} = 2^{a-b}.$$

The upper bound follows from

$$\prod_{i=0}^{b-1}(2^{a-i}-1) < \prod_{i=0}^{b-1}2^{a-i} = 2^{ab-b(b-1)/2}$$

and

$$\prod_{i=0}^{b-1}(2^{b-i}-1) = 2^{b(b+1)/2}\prod_{i=1}^{b}(1-2^{-i}) > 2^{b(b+1)/2}c. \quad \square$$

From Lemma 10 we get

$$c\,2^{r(k-h-r)} < \frac{\begin{bmatrix} h \\ k-r \end{bmatrix}\begin{bmatrix} n-h \\ r \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}}2^{(k-r)(n-h-r)} < c^{-2}\,2^{r(k-h-r)}$$

and so, for $h \geq k$ we have

$$c\sum_{r=1}^{\min(k,n-h)} r2^{r(k-h-r)} < k - E(e_h) < c^{-2}\sum_{r=1}^{\min(k,n-h)} r2^{r(k-h-r)}. \tag{6}$$

Similarly, for $h \leq k$ we have

$$c\sum_{r=1}^{\min(h,n-k)} r2^{r(h-k-r)} < h - E(e_h) < c^{-2}\sum_{r=1}^{\min(h,n-k)} r2^{r(h-k-r)}.$$

Remark: Let $\omega(k)$ be an integer valued function. Let $h = k+\omega(k)$ and $n \geq 2k+\omega(k)$. If $\omega(k) \to \infty$ when $k \to \infty$, then, by (6),

$$k - E(e_{k+\omega(k)}) \asymp 2^{-\omega(k)}.$$

We next consider the variance. First we need another lemma.

Lemma 11: Let P be a binary $k \times l$ matrix of rank s and Q a $k \times m$ matrix such that $(P \mid Q)$ has rank $s+t$. Then the number $\delta(k,p,s,t,u,v)$ of $k \times p$ matrices R such that $(P \mid R)$ has rank $s+u$ and $(P|Q|R)$ has rank $s+t+v$ is exactly

$$\delta(k,p,s,t,u,v) = 2^{s(p-v)}\begin{bmatrix} u \\ v \end{bmatrix}\begin{bmatrix} p \\ u \end{bmatrix}\prod_{l=1}^{u-v}(2^t-2^{l-1}) \times \prod_{j=s+t}^{s+t+v-1}(2^k-2^j).$$

Proof: As in the proof of Lemma 9, we may assume without loss of generality that

$$P = \begin{pmatrix} I_s \\ O_{t,s} \\ O_{k-s-t,s} \end{pmatrix} \qquad Q = \begin{pmatrix} O_{s,t} \\ I_t \\ O_{k-s-t,t} \end{pmatrix} \qquad R = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

where X is an arbitrary $s \times p$ matrix, Z is a $(k-s-t) \times p$ matrix of rank v, and Y is a $t \times p$ matrix such that $\binom{Y}{Z}$ has rank u. The matrix X can be chosen in 2^{sp} ways. By Lemma 8, Z can be chosen in $\left[\begin{smallmatrix} k-s-t \\ v \end{smallmatrix}\right]\left[\begin{smallmatrix} p \\ v \end{smallmatrix}\right][v]$ ways. For a given Z, by Lemma 9, Y can be chosen in

$$\gamma(p,t,v,u-v) = 2^{tv}\begin{bmatrix} t \\ u-v \end{bmatrix}\begin{bmatrix} p-v \\ u-v \end{bmatrix}[u-v]$$

ways. Combining and simplifying, the lemma follows. $\square$

Lemma 12: For all h we have

$$E(e_h^2) = \sum_{r=0}^{h}\sum_{s=0}^{h}\sum_{v=0}^{s-u}\sum_{u=0}^{\min(r,s)}\sum_{a=0}^{h} rs\frac{\binom{n}{a}\binom{n-a}{h-a}\binom{n-h}{h-a}}{\binom{n}{h}^2} \cdot 2^{u(2h-2a-r+u-v)+(r+v)(n-2h+a-k+r+v)} \cdot \frac{\begin{bmatrix} a \\ u \end{bmatrix}\begin{bmatrix} h-a \\ r-u \end{bmatrix}\begin{bmatrix} s-u \\ v \end{bmatrix}\begin{bmatrix} h-a \\ s-u \end{bmatrix}\begin{bmatrix} n-2h+a \\ k-r-v \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}} \cdot \prod_{l=1}^{s-u-v}(2^{r-u}-2^{l-1}).$$

Proof: There are

$$\binom{n}{a}\binom{n-a}{h-a}\binom{n-h}{h-a}$$

choices of h-subsets of $X, Y \subset \{1, 2, \cdots, n\}$ such that $|X \cap Y| = a$. The number of $k \times n$ matrices such that the submatrices corresponding to the columns with positions in $X \cap Y$, X, Y, $X \cup Y$ have rank u, r, s, and $r+v$, respectively, is

$$\sum_{v=0}^{s-u}\gamma(k,a,0,u)\gamma(k,h-a,u,r-u) \cdot \delta(k,h-a,u,r-u,s-u,v) \cdot \gamma(k,n-2h+a,r+v,k-r-v).$$

Hence

$$
\begin{aligned}
&E(e_h^2)\\
&= \frac{1}{\gamma(k,0,n,0,k)\binom{n}{h}^2} \sum_{r=0}^{h}\sum_{s=0}^{h}\sum_{v=0}^{s-u}\sum_{u=0}^{\min(r,s)}\sum_{a=0}^{h} rs\\
&\cdot \binom{n}{a}\binom{n-a}{h-a}\binom{n-h}{h-a}\\
&\cdot \sum_{v=0}^{s-u} \gamma(k,a,0,u)\gamma(k,h-a,u,r-u)\\
&\cdot \delta(k,h-a,u,r-u,s-u,v)\\
&\cdot \gamma(k,n-2h+a,r+v,k-r-v)\\
&= \sum_{r=0}^{h}\sum_{s=0}^{h}\sum_{v=0}^{s-u}\sum_{u=0}^{\min(r,s)}\sum_{a=0}^{h} rs \frac{\binom{n}{a}\binom{n-a}{h-a}\binom{n-h}{h-a}}{\binom{n}{h}^2}\\
&\cdot 2^{u(h-a-r+u)+u(h-a-v)+r(h-a-s+u)+(r+v)(n-2h+a-k+r+v)}\\
&\cdot \frac{\begin{bmatrix} a \\ u \end{bmatrix}\begin{bmatrix} h-a \\ r-u \end{bmatrix}\begin{bmatrix} s-u \\ v \end{bmatrix}\begin{bmatrix} h-a \\ s-u \end{bmatrix}\begin{bmatrix} n-2h+a \\ k-r-v \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}}\\
&\cdot \prod_{l=1}^{s-u-v} (2^{r-u}-2^{l-1}). \qquad \square
\end{aligned}
$$

Since $\mathrm{Var}\,(e_h) = E(e_h^2) - E(e_h)^2$, we can combine Theorem 10 and Lemma 12 to obtain $\mathrm{Var}\,(e_h)$. In general, it is a quite complicated expression.

As an example, we compute $\mathrm{Var}\,(e_{n-1})$. We note that we get a contribution to the sum only if $a = n-2$ or $a = n-1$. For $a = n-1$ we only get a contribution when $u = r = s \in \{k-1, k\}$ and $v = 0$. For $a = n-2$ we get contributions only for $k-2 \le u \le k$. We get

$$
\begin{aligned}
&E(e_{n-1}^2)\\
&= (k-1)^2 \frac{n-1}{n} \frac{\begin{bmatrix} n-2 \\ k-2 \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}} + (k-1)k\frac{n-1}{n}2^{k-1}\frac{\begin{bmatrix} n-2 \\ k-1 \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}}\\
&\quad + k(k-1)\frac{n-1}{n}2^{k-1}\frac{\begin{bmatrix} n-2 \\ k-1 \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}} + k^2\frac{n-1}{n}2^{k-1}\frac{\begin{bmatrix} n-2 \\ k-1 \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}}\\
&\quad + k^2\frac{n-1}{n}2^{2k}\frac{\begin{bmatrix} n-2 \\ k \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}}\\
&\quad + (k-1)^2\frac{1}{n}\frac{\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}} + k^2\frac{1}{n}2^k\frac{\begin{bmatrix} n-1 \\ k \end{bmatrix}}{\begin{bmatrix} n \\ k \end{bmatrix}}\\
&= (k-1)^2\frac{n-1}{n}\frac{(2^k-1)(2^{k-1}-1)}{(2^n-1)(2^{n-1}-1)}\\
&\quad + (3k^2-2k)\frac{n-1}{n}2^{k-1}\frac{(2^k-1)(2^{n-k}-1)}{(2^n-1)(2^{n-1}-1)}\\
&\quad + (k-1)^2\frac{1}{n}\frac{(2^k-1)}{(2^n-1)} + k^2\frac{1}{n}2^k\frac{(2^{n-k}-1)}{(2^n-1)}\\
&\quad + k^2\frac{n-1}{n}2^{2k}\frac{(2^{n-k}-1)(2^{n-k-1}-1)}{(2^n-1)(2^{n-1}-1)}.
\end{aligned}
$$

Simplifying and combining with (5) we get

$$
\mathrm{Var}\,(e_{n-1}) = \frac{(2^k-1)(2^{n-1}-2^{k-1})(2^n-n-1)}{n(2^n-1)^2(2^{n-1}-1)}.
$$

Taking $\epsilon = \frac{2^k-1}{2^n-1}$ in Tchebychev's inequality we get

$$
P\left(e_{n-1} \le k - 2\frac{2^k-1}{2^n-1}\right) \le \frac{(2^{n-1}-2^{k-1})(2^n-n-1)}{n(2^k-1)(2^{n-1}-1)}.
$$

In particular, if $n \to \infty$ and $n-k = o(\log n)$, then

$$
P\left(e_{n-1} \le k - 2\frac{2^k-1}{2^n-1}\right) \to 0.
$$

Acknowledgment

The authors wish thank one of the anonymous referees who pointed out that the proofs of Lemmas 8, 9, and 11 could be simplified and who sketched the proof of Lemmas 8 and 9 given above.

References

[1] L. Carlitz, "Note on a paper of Laksov," *Math. Scand.*, vol. 19, pp. 38–40, 1966.

[2] G. C. Clark and J. B. Cain, *Error-Correcting Coding for Digital Communication*. New York: Plenum, 1981.

[3] G. D. Forney, "Dimension/length profiles and trellis complexity of linear block codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1741–1752, 1994.

[4] T. Helleseth, T. Kløve, V. Levenshtein, and Ø. Ytrehus, "Bounds on the minimum support weights," *IEEE Trans. Inform. Theory*, vol. 41, pp. 432–440, 1995.

[5] T. Helleseth, T. Kløve, and J. Mykkeltveit, "The weight distribution of irreducible cyclic codes with block lengths $n_1((q^l-1)/N)$," *Discr. Math.*, vol. 18, pp. 179–211, 1977.

[6] T. Kasami, T. Takata, T. Fujiwara, and S. Lin, "On the optimum bit orders with respect to the state complexity of trellis diagrams for binary linear codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 242–245, 1993.

[7] T. Kløve, "The weight distribution of linear codes over GF(q^l) having generator matrix over GF(q)," *Discr. Math.*, vol. 23, pp. 159–168, 1978.

[8] D. Laksov, "Linear recurring sequences over finite fields," *Math. Scand.*, vol. 16, pp. 181–196, 1965.

[9] V. Levenshtein, "Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1303–1321, 1995.

[10] J. Simonis, "The effective length of subcodes," *Applicable Algebra in Eng., Commun. and Computing*, vol. 5, pp. 371–377, 1994.

[11] A. Vardy and Y. Be'ery, "Maximum-likelihood soft decision decoding of BCH codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 546–554, 1994.

[12] V. K. Wei, "Generalized Hamming weights for linear codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1412–1418, 1991.