Estimates of the Bit Error Probabilities for Linear Block Codes and Symmetric Binary–Input Memoryless Channels

Vladimir B. Balakirsky
EIDMA, c/o TU/e
P.O.Box 513, 5600 MB Eindhoven
The Netherlands
Email : v.b.balakirsky@tue.nl

Abstract — **A coset representation of the bit error probabilities for binary linear block codes and arbitrary symmetric memoryless channels is derived. Such a representation leads to simple upper bounds on the bit error probabilities for code components expressed via $2n$ code spectra, where n is the code length, and to a "bounded distance decoding" algorithm. The performance of the algorithm is illustrated for the $(23, 12)$ Golay code and a particular channel having non–binary output alphabet where it attains a sub–optimum performance.**

I. Introduction

We address the decoding problem when the decoder has to form an estimate that minimizes the bit error probabilities for code components. This problem was under considerations in many papers and we refer to the introductory part of [1], which gives a good overview of current understanding of the problem.

Our results can be summarized as follows.

(1) Interpretation of data transmission over an arbitrary symmetric memoryless channel as transmission over a time–varying binary symmetric channel with side information at the decoder leads to extension of the known approaches for binary symmetric channels [2], [3] to arbitrary binary–input symmetric channels.

(2) We derive an upper bound on the bit error probabilities for code components. This bound is expressed via code spectra in such a way that at most $2n$ spectra are needed for computations, where n denotes the code length. If the encoder is systematic, then we simultaneously obtain an upper bound on the bit error probabilities for information components.

(3) The upper bound on the bit error probabilities for code components gives the exact value of these parameters attained with the decoding algorithm, which is similar to "the bounded distance decoding algorithm" developed for minimizing the word error probability. The algorithm is illustrated for the $(23, 12)$ Golay code.

Decoding for minimizing the bit error probability is usually considered as a problem, which is more difficult than decoding for minimizing the word error probability because the optimum decoding rule is expressed via sums of conditional probabilities of the received vector for all codewords. This rule can be simplified for linear codes and binary symmetric channels where direct computation requires inspection of 2^{n-k} cosets. We concentrate on the bit error probabilities for code components and arrive at the point that bounded distance decoding, where we need to investigate only error patterns of the Hamming weight at most $t_{\min} \triangleq \lfloor (d_{\min} - 1)/2 \rfloor$, can achieve the optimum performance ($d_{\min}$ denotes the minimum distance of the code). Examples, for which this statement is true, include the $(6, 3)$ shortened Hamming code, the $(10, 5)$ van Gils code, and the $(23, 12)$ Golay code. Our decoding algorithm is universal in a sense that, for any code, one can find the threshold values $\varepsilon_1, \ldots, \varepsilon_{t_{\min}}$ and run the procedure described for the $(23, 12)$ Golay code.

We estimate the bit error probability for the j–th code component from above using 2 code spectra : the spectrum of codewords having 0 at the j–th position and the spectrum of codewords having 1 at the j–th position. This bound gives exact value of the bit error probability for the j–th code component when the decoder does not take into account the standard array of the code and processes only error vectors of the Hamming weight $0, 1, \ldots, t_{\min}$.

We also show that decoding problem for arbitrary symmetric channels having binary input alphabet and $(2L)$–output alphabet can be reduced to the decoding problem for a time–varying binary symmetric channel whose crossover probabilities at each time instant are independently chosen from a certain set containing L entries. As a result, previous conclusions for binary symmetric channels can be extended to arbitrary binary–input symmetric memoryless channels.

II. Notations and basic ideas

A binary linear block (n, k)–code is understood as a collection of 2^k binary vectors of length n. The information vector $\mathbf{u} = (u_1, \ldots, u_k) \in \{0, 1\}^k$ is mapped to the codeword $\mathbf{uG} \triangleq ((\mathbf{uG})_1, \ldots, (\mathbf{uG})_n)$, where $\mathbf{G}$ is the $k \times n$ generator matrix of the encoder. To simplify notations, we will say "the encoder $\mathbf{G}$". Let $\mathcal{E}^n$ denote some set of representatives of cosets $\bigcup_{\mathbf{z}\in\{0,1\}^k} \{\mathbf{zG} \oplus \mathbf{e}\}$ for the encoder $\mathbf{G}$.

Suppose that there is a discrete memoryless channel defined by the conditional probability distributions $(W_x(y), y \in \mathcal{Y} \triangleq \{0,\ldots,2L-1\}), x = 0,1$, in a sense that

$$W(\mathbf{y}|\mathbf{x}) \triangleq \prod_{j=1}^{n} W_{x_j}(y_j)$$

is the conditional probability to receive the vector $\mathbf{y} \in \mathcal{Y}^n$ when the vector $\mathbf{x} \in \{0,1\}^n$ was sent. We will assume that the $2 \times (2L)$ matrix of crossover probabilities has the form

$$\mathbf{W} = \begin{bmatrix} q(0) & \ldots & q(L-1) & p(L-1) & \ldots & p(0) \\ p(0) & \ldots & p(L-1) & q(L-1) & \ldots & q(0) \end{bmatrix}$$

where $q(y) > p(y) \geq 0$ for all $y \in \{0,\ldots,L-1\}$ and $\sum_{y=0}^{L-1}(p(y)+q(y)) = 1$. We will also denote

$$\mathcal{P} \triangleq \bigcup_{y=0}^{L-1} \Big\{ p(y)/(p(y)+q(y)) \Big\}$$

and assume that $|\mathcal{P}| = L$. Channels having properties above will be referred to as *symmetric binary–input memoryless channels.* If $L = 1$, then $\mathbf{W}$ determines a binary symmetric channel having the crossover probability $p \triangleq p(0)$, which will be referred to as a $BSC(p)$.

For all $\mathbf{y} \in \mathcal{Y}^n$, let $\varphi'(\mathbf{y}) \triangleq (\varphi'_1(\mathbf{y}),\ldots,\varphi'_n(\mathbf{y})) \in \{0,1\}^n$ be the estimate formed at the output of the decoder (note that $\varphi'(\mathbf{y})$ can be not a codeword). Then

$$P'_{\mathrm{b},j} \triangleq \min_{\{\varphi'(\mathbf{y})\}} \sum_{\mathbf{u}\in\{0,1\}^k} \sum_{\mathbf{y}:\varphi'_j(\mathbf{y})\neq(\mathbf{uG})_j} \frac{1}{2^k} W(\mathbf{y}|\mathbf{uG}) \qquad (1)$$

is called *the bit error probability for the j–th code component.* If the encoder is systematic in a sense that $(\mathbf{uG})_j = u_j$ for all $j = 1,\ldots,k$, then $P'_{\mathrm{b},1},\ldots,P'_{\mathrm{b},k}$ are also the bit error probabilities for information components.

Let us introduce the following notations.

(a) *For all $y \in \{0,\ldots,2L-1\}$, let*

$$\xi(y) \triangleq \frac{\min\{W_0(y),W_1(y)\}}{W_0(y)+W_1(y)} = \frac{p(y)}{p(y)+q(y)} \qquad (2)$$

(b) *For all $y \in \{0,\ldots,L-1\}$, let*

$$\pi(\xi)\Big|_{\xi=\xi(y)} \triangleq W_0(y)+W_1(y) = p(y)+q(y).$$

Notice that

$$\sum_{\xi\in\mathcal{P}} \pi(\xi)\xi = \sum_{y=0}^{L-1} p(y).$$

(c) *For all $\xi \in \mathcal{P}$ and $x \in \{0,1\}$, let*

$$V_\xi(x) \triangleq |(x \oplus 1) - \xi|.$$

(d) *For all $\boldsymbol{\xi} \in \mathcal{P}^n$ and $\mathbf{x} \in \{0,1\}^n$, let*

$$\Pi(\boldsymbol{\xi}) \triangleq \prod_{j=1}^{n} \pi(\xi_j), \quad V(\mathbf{x}|\boldsymbol{\xi}) \triangleq \prod_{j=1}^{n} V_{\xi_j}(x_j).$$

Example. *Let $L = 2$ and*

$$\mathbf{W} = \begin{bmatrix} 0.70 & 0.20 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.20 & 0.70 \end{bmatrix}. \qquad (3)$$

Then $(\xi(0),\ldots,\xi(3)) = (1/15,1/5,1/5,1/15)$, $\mathcal{P} = \{1/15,1/5\}$, *and* $(\pi(1/15),\pi(1/5)) = (3/4,1/4)$.

Our considerations for arbitrary symmetric channels are based on the following observations that can be explained for the channel defined in (3). Let us introduce the data transmission system where the j–th code bit is transmitted either over a $BSC(1/15)$, or over a $BSC(1/5)$ (see Figure 1). The selection is made by a jammer who chooses the first channel with probability 3/4 and the second channel with probability 1/4. The jammer also informs the decoder about his choice by sending him $\xi_j \in \{1/15,1/5\}$. These operations are independently repeated for all $j = 1,\ldots,n$. The decoder has to construct an estimate of the transmitted codeword for a given pair of vectors $(\hat{\mathbf{y}},\boldsymbol{\xi}) = ((\hat{y}_1,\xi_1),\ldots,(\hat{y}_n,\xi_n))$, where $\hat{y}_j$ denotes the j–th received bit. It is easy to see that this problem coincides with the original problem of estimating the transmitted codeword for a given vector $\mathbf{y} = (y_1,\ldots,y_n)$ received at the output of the channel having the matrix of crossover probabilities (3), since there are component–wise one–to–one mappings $0 \leftrightarrow (0,1/15)$, $1 \leftrightarrow (0,1/5)$, $2 \leftrightarrow (1,1/5)$, $3 \leftrightarrow (1,1/15)$ that translate $\mathbf{y}$ to the pair $(\hat{\mathbf{y}},\boldsymbol{\xi})$. Such a translation to data transmission over a time–varying binary symmetric channel makes sense from technical point of view. Namely, let us express the probability $P'_{\mathrm{b},j}$ as

$$P'_{\mathrm{b},j} = \sum_{\boldsymbol{\xi}\in\mathcal{P}^n} \Pi(\boldsymbol{\xi}) P'_{\mathrm{b},j}(\boldsymbol{\xi}) \qquad (4)$$

where $P'_{\mathrm{b},j}(\boldsymbol{\xi})$ is the bit error probability for the j–th code component when the 1–st code bit is transmitted over a $BSC(\xi_1)$, ..., the n–th code bit is transmitted over a $BSC(\xi_n)$. We fix a vector $\boldsymbol{\xi} \in \mathcal{P}^n$ and use the decoding methods and their analysis for a BSC to evaluate the probability $P'_{\mathrm{b},j}(\boldsymbol{\xi})$. By (4), the final result is obtained as a weighted sum of the results for all possible vectors $\boldsymbol{\xi}$, where the weight is defined as $\Pi(\boldsymbol{\xi})$.

III. Formulation of the results and the decoding algorithms

Theorem.

1. *Given a received vector $\mathbf{y} \in \mathcal{Y}^n$, let the pair of vectors $(\mathbf{z},\mathbf{e}) \in \{0,1\}^k \times \mathcal{E}^n$ be defined by the equation*

$$\hat{\mathbf{y}} = \mathbf{zG} \oplus \mathbf{e}$$

where

$$\hat{y}_j \triangleq \begin{cases} 0, & \textit{if } y_j \in \{0,\ldots,L-1\} \\ 1, & \textit{if } y_j \in \{L,\ldots,2L-1\} \end{cases} \qquad (5)$$

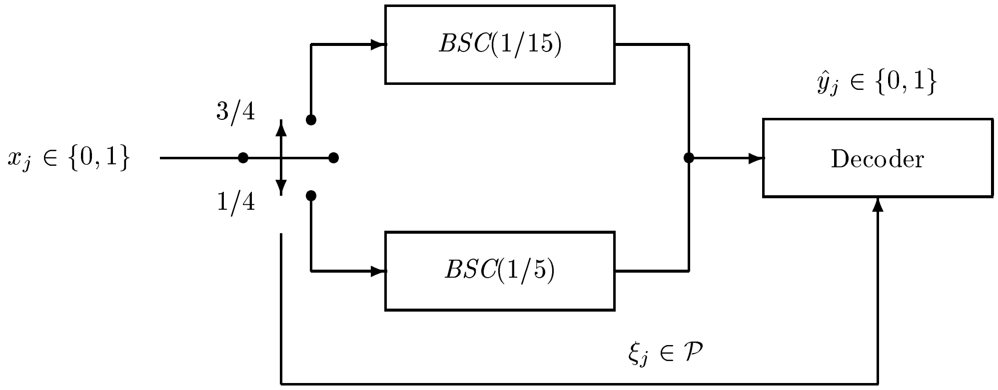

Figure 1: The scheme, which is equivalent to data transmission over a channel defined in (3).

and let $\xi(\mathbf{y}) = (\xi(y_1), \ldots, \xi(y_n))$ *be the vector whose components are defined by* (2). *Then, for all* $j = 1, \ldots, n$,

$$\varphi_j^{(\mathrm{opt})\prime}(\mathbf{y}) = \arg\max_{x=0,1} \upsilon_j^{(x\oplus\hat{y}_j\oplus e_j)\prime}(\mathbf{e}|\xi(\mathbf{y}))$$

minimizes the sum on the right–hand side of (1) *and*

$$P'_{\mathrm{b},j} = \sum_{\boldsymbol{\xi}\in\mathcal{P}^n} \Pi(\boldsymbol{\xi}) \sum_{\mathbf{e}\in\mathcal{E}^n} \min\Big\{\upsilon_j^{(0)\prime}(\mathbf{e}|\boldsymbol{\xi}), \upsilon_j^{(1)\prime}(\mathbf{e}|\boldsymbol{\xi})\Big\}$$

where

$$\upsilon_j^{(x)\prime}(\mathbf{e}|\boldsymbol{\xi}) \triangleq \sum_{\mathbf{u}:(\mathbf{uG})_j=x} V(\mathbf{uG}\oplus\mathbf{e}|\boldsymbol{\xi}), \quad x = 0,1.$$

2. *For* $x = 0,1$ *and* $d \in \{0, \ldots, n\}$, *let*

$$M_j^{(x)\prime}(d) \triangleq \Big|\Big\{\mathbf{u} : (\mathbf{uG})_j = x,\ \mathrm{wt}(\mathbf{uG}) = d\Big\}\Big|$$

denote the number of codeword of weight d *whose* j*–th component is equal to* x. *Then, for all* $j = 1, \ldots, n$,

$$\tilde{p} - P'_{\mathrm{b},j} \ge \sum_{\xi\in\mathcal{P}} \pi(\xi) \sum_{t=1}^{t_{\min}} \sum_{e=0,1} \max\{\, \Delta_{j,t}^{(e)}(\xi,\tilde{p}),\, 0\,\} \tag{6}$$

where

$$\begin{aligned} \tilde{p} &\triangleq \sum_{\xi\in\mathcal{P}} \pi(\xi)\xi \\ \Delta_{j,t}^{(e)}(\xi,\tilde{p}) &\triangleq \sum_{d=0}^{n} \Big[\, \xi M_j^{(0)\prime}(d)\Upsilon_{t-1}^{(d)}(\tilde{p}) \\ &\quad - (1-\xi)M_j^{(1)\prime}(d)\Upsilon_{t-1}^{(d-1)}(\tilde{p})\,\Big] \quad (7) \\ \Upsilon_t^{(d)}(\tilde{p}) &\triangleq (1-\tilde{p})^{n-1} \\ &\quad \cdot \sum_{\tau=0}^{t} \binom{d}{\tau}\binom{n-1-d}{t-\tau}\left(\frac{\tilde{p}}{1-\tilde{p}}\right)^{d+t-2\tau}. \end{aligned}$$

Let us illustrate considerations for the $(23,12)$ Golay code whose spectra are given in Table 1. If $L = 1$ and $p = p(0)$ (i.e., we deal with the $BSC(p)$), then (6) brings the inequality

Table 1: The spectra of the $(23,12)$ Golay code, $j = 1, \ldots, 23$.

d	$M_j^{(0)\prime}(d)$	$M_j^{(1)\prime}(d)$
0	1	0
7	176	77
8	330	176
11	672	616
12	616	672
14	176	330
15	77	176
23	0	1

$$p - P'_{\mathrm{b},j} \ge \sum_{t=1}^{3} \max\{\, \Delta_{j,t}^{(1)}(p,p),\, 0\,\}, \quad j = 1, \ldots, 23$$

and $\Delta_{j,t}^{(1)}(p,p) > 0$ if and only if $p \in (0, \varepsilon_t)$, where

$$(\varepsilon_1, \varepsilon_2, \varepsilon_3) \approx (0.279, 0.228, 0.117)$$

are found using (7). Thus, the algorithm, for which this inequality is tight, is as follows.

- $p \in (0, \varepsilon_3)$: *Find the vector* $\mathbf{e}$ *of the smallest Hamming weight such that* $\mathbf{y}\oplus\mathbf{e}$ *is a codeword and set* $\varphi'(\mathbf{y}) = \mathbf{y}\oplus\mathbf{e}$.
- $p \in (\varepsilon_3, \varepsilon_2)$: *If* $\mathbf{y}\oplus\mathbf{e}$ *is a codeword for some vector* $\mathbf{e}$ *with* $\mathrm{wt}(\mathbf{e}) \le 2$, *then set* $\varphi'(\mathbf{y}) = \mathbf{y}\oplus\mathbf{e}$. *Otherwise, set* $\varphi'(\mathbf{y}) = \mathbf{y}$.
- $p \in (\varepsilon_2, \varepsilon_1)$: *If* $\mathbf{y}\oplus\mathbf{e}$ *is a codeword for some vector* $\mathbf{e}$ *with* $\mathrm{wt}(\mathbf{e}) \le 1$, *then set* $\varphi'(\mathbf{y}) = \mathbf{y}\oplus\mathbf{e}$. *Otherwise, set* $\varphi'(\mathbf{y}) = \mathbf{y}$.
- $p \in (\varepsilon_1, 1/2)$: *Set* $\varphi'(\mathbf{y}) = \mathbf{y}$.

Notice that the algorithm above differs from the maximum likelihood decoding : $\varphi^{(\mathrm{ML})}(\mathbf{y}) = \mathbf{y}\oplus\mathbf{e}$, where $\mathbf{e}$ is the vector of the smallest Hamming weight such that $\mathbf{y}\oplus\mathbf{e}$ is a codeword. The performance of both algorithms

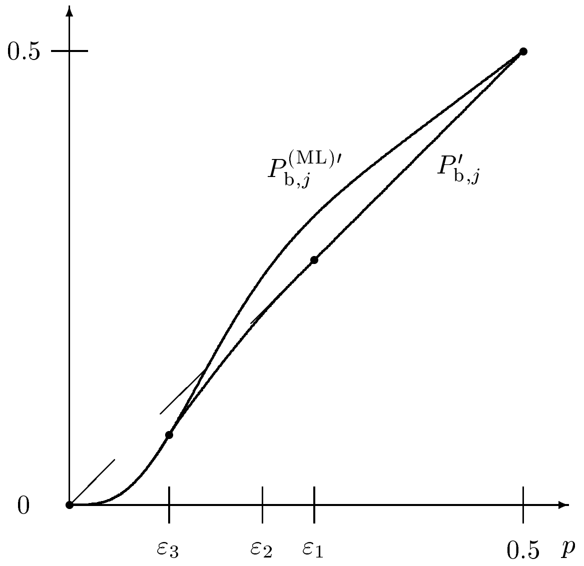

Figure 2: *BSC*(p) and the $(23,12)$ Golay code : the bit error probabilities for the optimum decoding $P'_{\mathrm{b},j}$ and for the maximum likelihood decoding $P^{(\mathrm{ML})\prime}_{\mathrm{b},j}$.

is illustrated in Figure 2. Furthermore, the analysis of the standard array of the $(23,12)$ shows that our algorithm is optimal. Similar conclusion can be also derived for the shortened Hamming $(6,3)$–code, for the van Gils $(10,5)$–code, and for some other codes, which are not perfect.

The algorithm can be also extended to binary–input symmetric channels with a non–binary output alphabet. For all $j = 1,\ldots,n$, $w \in \{0,\ldots,t_{\min}\}$, and $\xi \in \mathcal{P}$, let $\varepsilon_{j,w}(\xi)$ denote the largest value of $p \in (0, 0.5]$ such that $\Delta^{(1)}_{j,w}(\xi, p) > 0$ and let $\varepsilon_{j,w}(\xi) = 0$ for all $w > t_{\min}$. In particular, for the $(23,12)$–code,

$$\begin{bmatrix} \varepsilon_{j,0}(1/5) & \varepsilon_{j,0}(1/15) \\ \varepsilon_{j,1}(1/5) & \varepsilon_{j,1}(1/15) \\ \varepsilon_{j,2}(1/5) & \varepsilon_{j,2}(1/15) \\ \varepsilon_{j,3}(1/5) & \varepsilon_{j,3}(1/15) \end{bmatrix} \approx \begin{bmatrix} 0.5 & 0.5 \\ 0.261 & 0.223 \\ 0.220 & 0.174 \\ 0.148 & 0.092 \end{bmatrix}. \quad (8)$$

Given a received vector $\mathbf{y} \in \mathcal{Y}^n$, let $\mathbf{e}$ be the binary vector of the smallest Hamming weight such that $\hat{\mathbf{y}} \oplus \mathbf{e}$ is a codeword, where $\hat{\mathbf{y}}$ is the vector defined in (5), and let $w = \mathrm{wt}(\mathbf{e})$.

- ⊙ *For all* $j = 1,\ldots,n$, *set* $\varphi'_j(\mathbf{y}) = \hat{y}_j \oplus 1$ *if* $e_j = 1$ *and* $\tilde{p} \le \varepsilon_{j,w}(\xi_j)$; *otherwise, set* $\varphi'_j(\mathbf{y}) = \hat{y}_j$.

If the channel is defined by (3), then $\tilde{p} = 0.1$ and (8) implies that our algorithm outputs $\hat{y}_j \oplus e_j$ at the j–th position, except for the case $e_j = 1$ and $\xi_j = 1/15$ where $\varphi'_j(\mathbf{y}) = \hat{y}_j$.

The performance of the algorithm above is very close to the performance of the optimal decoding algorithm for minimizing the bit error probability for the channel (3), and we are convinced by some other examples of channels and linear block codes that such a conclusion

Table 2: The values of $y^{(t)}, t = 1,2,3$, for the $(23,12)$ Golay code, $\sigma = 1/\sqrt{2}$.

E	$\tilde{p}$	$y^{(1)}/\sqrt{E}$	$y^{(2)}/\sqrt{E}$	$y^{(3)}/\sqrt{E}$
0.5	0.158	5.28	3.13	1.20
1.0	0.078	5.12	3.29	1.51
1.5	0.041	4.79	3.12	1.51
2.0	0.022	4.53	2.99	1.45
2.5	0.012	4.35	2.86	1.40
3.0	0.007	4.20	2.77	1.37

is typical. One of these examples is data transmission over the AWGN channel when 0's and 1's are sent by the signals having the amplitudes $\pm\sqrt{E}$ and the Golay code is used for encoding. One can easily extend the result of the theorem to the continuous symmetric channels by assuming sufficiently many levels of quantization of the received signal. In this case, if y is the amplitude of the signal received at the j–th position, then

$$\xi(y) = \frac{e^{-\frac{1}{2\sigma^2}\max\{(y-\sqrt{E})^2,(y+\sqrt{E})^2\}}}{e^{-\frac{(y-\sqrt{E})^2}{2\sigma^2}} + e^{-\frac{(y+\sqrt{E})^2}{2\sigma^2}}}$$

$$\tilde{p} = \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\sqrt{E})^2}{2\sigma^2}} dz$$

where σ^2 denotes the variance of the noise. We can form a binary vector $\hat{\mathbf{y}}$ by quantizing the components of the received vector $\mathbf{y}$ and use the following generalized version of the algorithm described for a *BSC*.

- ⊙ *Find the vector* $\mathbf{e}$ *of the smallest Hamming weight such that* $\hat{\mathbf{y}} \oplus \mathbf{e}$ *is a codeword and set* $\varphi'(\mathbf{y}) = \hat{\mathbf{y}}$.
- ⊙ *For all* $j = 1,\ldots,23$ *such that* $e_j = 1$, *check whether* $\Delta^{(1)}(\xi(y_j), \tilde{p}) > 0$ *or not. If the answer is positive, then invert the* j*–th component of the vector* $\varphi'(\mathbf{y})$.

In Table 2 we show the values of $y^{(t)}$, $t = 1,2,3$ defined as the largest amplitudes $|y|$ such that $\Delta^{(1)}(\xi(y^{(t)}), \tilde{p}) > 0$ when $\sigma = 1/\sqrt{2}$.

References

[1] A. Abedi, P. Chaudhari, and A. K. Khandani, "On some properties of the bit decoding algorithms," in : *Proc. 7–th Canadian Workshop on Information Theory*, Vancouver, British Columbia, pp. 106–109, June 2001.

[2] A. B. Kiely, J. T. Coffey, and M. R. Bell, "Optimal information bit decoding of linear block codes," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 130–140, 1995.

[3] S. Coffey and H. Chen, "Bit error probability and encoder structure," in : *Coding, Communication and Broadcasting*, P. Farrel, M. Darnell, B. Honary (eds.), Research Studies Press Ltd., Baldock, Herdfordshire, England, pp. 29–38, 2000.