On Two Probabilistic Decoding Algorithms for Binary Linear Codes

Miodrag Živković

[*]

Abstract

A generalization of Sullivan inequality on the ratio of the probability of a linear code to that of any of its cosets is proved. Starting from this inequality, a sufficient condition for successful decoding of linear codes by a probabilistic method is derived. A probabilistic decoding algorithm for "low–density parity–check codes" is also analyzed. The results obtained allow one to estimate experimentally the probability of successful decoding using these probabilistic algorithms.

Index Terms — linear code, inequality, probabilistic decoding, low–density parity–check code

[*]The author is with the Institute of Applied Mathematics and Electronics, Belgrade, Yugoslavia

Introduction

In this paper we shall prove a generalization of Sullivan [11] inequality on the ratio of the probability of a subgroup (linear subspace, linear code) of the additive group of the field $\mathrm{GF}(2^n)$ (linear space over the field $\mathrm{GF}(2)$) to that of any of its cosets. The generalized inequality gives a lower bound on the ratio of the probability of a subgroup of the additive group of $\mathrm{GF}(q^n)$ (q is a power of a prime) to that of an arbitrary coset of it, for a more general probability distribution on the set $\mathrm{GF}(q^n)$.

We shall perform an analysis of a probabilistic decoding method of linear codes based on the special case of this inequality for $q = 2$. The algorithm can be described as follows (see [4, pp. 152], [1], [3], [8, Algorithm B]). For a given binary linear (n, k) code one has to define a mapping (depending on the received message) of a vector of *a priori* error probabilities into a vector of *a posteriori* ones, using a different set of parity checks for each bit of the received message. This mapping is first applied to the n–dimensional vector whose coordinates are equal to the error probabilities of a DMC (discrete memoryless channel), then to the result of the mapping, and so on. After a number of iterations, the obtained vector is used to correct errors in the received message. A sufficient condition is given for convergence of the vector sequence of "*a posteriori* error probabilities" to the error vector, and equivalently for successful decoding.

This method of analysis enables one to estimate experimentally the probability of successful decoding for any given linear code and for any chosen family of parity–check sets. By the example of a linear $(512, 100)$ code we shall illustrate the dependence of the successful decoding probability on the channel noise level. As it is known, such probabilistic algorithms are very efficient compared with other general decoding algorithms for linear codes (see [6], [8]).

The second algorithm that we shall consider is similar to Gallager probabilistic decoding method of "low–density parity–check codes" [6], and it is applicable only to these codes. An analysis of a probabilistic decoding method (of these codes) can be found in [6], where an average error probability is treated.

1 A generalization of an inequality on the ratio of the probability of a linear code to that of any of its cosets

Some necessary notation is introduced, and a generalization of Sullivan inequality (Theorem 1) is given. The proof of this inequality, similar to the proof of Sullivan inequality in [11], is given in the Appendix.

Let q be a power of a prime, and let V_n denote the additive group of the field $\mathrm{GF}(q^n)$. The elements of the set V_n are the vectors $\underline{\mathbf{v}} = [v_1\ v_2\ \ldots\ v_n]^T$, where $v_i \in V_1$ for $1 \leq i \leq n$, and T denotes transposition. A probability distribution $P(\cdot)$ on the set V_n is defined, so that for any $A \subset V_n$ we have

$$P(A) = \sum_{\underline{\mathbf{v}} \in A} \prod_{i=1}^{n} p_{i,v_i}. \tag{1}$$

Here $p_{i,j},\ 1 \leq i \leq n,\ j \in V_1$, are real, non-negative numbers, satisfying the conditions

$$\begin{cases} p_{i,j} = p_i, \quad j \in V_1,\ j \neq 0 \\ p_{i,0} = 1 - (q-1)p_i \end{cases}, \quad 0 \leq p_i < 1/q, \qquad 1 \leq i \leq n. \tag{2}$$

Let G be an arbitrary subgroup of order q^k of the group $V_n,\ 1 \le k < n$ (q–ary linear code), and let K be an arbitrary coset of G. In this paper a lower bound on the ratio $P(G)/P(K)$ is given. This problem is proposed and solved in [11] for $q = 2$ and $0 < p_i = p < 1/2,\ 1 \le i \le n$. Then the following inequality holds

$$\frac{P(G)}{P(K)} \ge \frac{1 + (1-2p)^{k+1}}{1 - (1-2p)^{k+1}} > 1. \tag{3}$$

Another proof of this inequality is given in [10], and some of its applications are listed.

Let $\underline{\mathbf{j}} = (j_1, j_2, \ldots, j_n)$ be an arbitrary permutation of the set of indices $\{1, 2, \ldots, n\}$, k an integer, $0 \le k < n$, and $u \in V_1$. Denote by $C_{\underline{\mathbf{j}},k,u}$ the following subset of V_n

$$C_{\underline{\mathbf{j}},k,u} = \left\{ \underline{\mathbf{v}} \in V_n \mid v_{j_1} + \cdots + v_{j_{k+1}} = u,\ v_{j_{k+2}} = \cdots = v_{j_n} = 0 \right\}. \tag{4}$$

The set $C_{\underline{\mathbf{j}},k,0}$ is a subgroup of order q^k, and the sets $C_{\underline{\mathbf{j}},k,u},\ u \ne 0$, are the cosets of this subgroup. It is easily seen that for $q = 2$ the lower bound on the ratio $P(G)/P(K)$ in (3) is equal to the ratio $P(C_{\underline{\mathbf{j}},k,0})/P(C_{\underline{\mathbf{j}},k,1})$.

Consider the ratio $P(C_{\underline{\mathbf{j}},k,0})/P(C_{\underline{\mathbf{j}},k,u}),\ u \in V_1,\ u \ne 0$ in a more general case, where q is an arbitrary power of a prime. Then we have

$$\begin{aligned} \frac{P(C_{\underline{\mathbf{j}},k,0})}{P(C_{\underline{\mathbf{j}},k,u})} &= \frac{1 + (q-1)\prod_{i=1}^{k+1}(1 - qp_{j_i})}{1 - \prod_{i=1}^{k+1}(1 - qp_{j_i})} \\ &= 1 - q + \frac{q}{1 - \prod_{i=1}^{k+1}(1 - qp_{j_i})} \ . \end{aligned}$$

Similarity of this expression to the lower bound in (3) suggests to introduce the function $F_k(\underline{\mathbf{p}})$ by

$$F_k(\underline{\mathbf{p}}) = \frac{1 + (q-1)\prod_{i=1}^{k+1}(1 - qp_{l_i})}{1 - \prod_{i=1}^{k+1}(1 - qp_{l_i})}, \tag{5}$$

where the permutation $\underline{\mathbf{l}}$ is such that

$$p_{l_1} \geq p_{l_2} \geq \cdots \geq p_{l_n}. \tag{6}$$

Then for any $k,\ 1 \leq k < n,$ we have the inequality

$$P(C_{\underline{\mathbf{j}},k,0})/P(C_{\underline{\mathbf{j}},k,u}) \geq P(C_{\underline{\mathbf{l}},k,0})/P(C_{\underline{\mathbf{l}},k,u}) = F_k(\underline{\mathbf{p}}).$$

The expression of $F_k(\underline{\mathbf{p}})$ depends on the $k+1$ largest coordinates of the vector $\underline{\mathbf{p}} = (p_1, p_2, \ldots, p_n)$. For any $k,\ 0 \leq k < n$, the function $F_k(\underline{\mathbf{p}})$ has the following properties (which can easily be verified).

Property 1. $F_k(\underline{\mathbf{p}}) > 1$.

Property 2. $F_{k+1}(\underline{\mathbf{p}}) \leq F_k(\underline{\mathbf{p}}), \quad k < n-1$.

Property 3. $F_k(\underline{\mathbf{p}})$ is a non-increasing function of every coordinate of the vector $\underline{\mathbf{p}}$.

Concerning the lower bound on the probability ratio $P(C_{\underline{\mathbf{j}},k,0})/P(C_{\underline{\mathbf{j}},k,u})$, $u \neq 0$, the question arises, whether it is equal to the lower bound on the ratio $P(G)/P(K)$ when G is an arbitrary subgroup of order q^k of the group V_n, K an arbitrary coset of it, and the probability distribution $P(\cdot)$ satisfies constraint (2). An affirmative answer is given by the following theorem, a generalization of inequality (3) in [11].

Theorem 1 *Suppose that the probability distribution over V_n is given by (1), where the parameters $p_{i,j}$, $1 \leq i \leq n$, $j \in V_1$, satisfy constraint (2). If G is an arbitrary subgroup of order q^k, $0 \leq k < n$, of V_n and if K is an arbitrary proper coset of G, then the following inequality holds*

$$P(G)/P(K) \geq F_k(\underline{\mathbf{p}}) > 1, \tag{7}$$

where the function $F_k(\underline{\mathbf{p}})$ is defined by (5). The ratio $P(G)/P(K)$ reaches the lower bound in this inequality if $G = C_{\underline{\mathbf{l}},k,0}$ and $K = C_{\underline{\mathbf{l}},k,u}$, $u \neq 0$, where $\underline{\mathbf{l}}$ is a permutation of indices satisfying (6). □

The proof is carried out by induction over the order of G, and for fixed order of G by induction over the coset leader weight of K (see the Appendix).

Let us briefly consider the case where the probability distribution $P(\cdot)$ on V_n satisfies the more general constraint

$$\begin{cases} \sum_{j=0}^{q-1} p_{i,j} = 1 \\ p_{i,0} \geq p_{i,j}, \quad j \in V_1, \ j \neq 0 \end{cases}, \qquad 1 \leq i \leq n. \tag{8}$$

Finding a lower bound on the ratio $P(G)/P(K)$ is much harder in this case. Even more, for a fixed $u \in V_1$, $u \neq 0$, if G is a subgroup $C_{\underline{\mathbf{j}},k,0}$ of type (4), and if K is its corresponding coset $C_{\underline{\mathbf{j}},k,u}$, then it is hard to find the permutation $\underline{\mathbf{j}}$ of indices for which $P(G)/P(K)$ reaches its lowest value. This fact can be illustrated by the following example.

Example 1 Let $q = 3$, $n = 4$, $k = 1$, and let the parameters $p_{1,0}$, $p_{1,1}$, $p_{1,2}$; $p_{2,0}, \ldots, p_{4,2}$ have the following values 0.6, 0.3, 0.1; 0.6, 0.1, 0.3; 0.6, 0.1, 0.3; 0.7, 0.2, 0.1, satisfying (8). The probability ratio

$$\begin{aligned} R(j_1, j_2) &= P(C_{\underline{\mathbf{j}},k,0})/P(C_{\underline{\mathbf{j}},k,1}) \\ &= \frac{p_{j_1,0}p_{j_2,0} + p_{j_1,1}p_{j_2,2} + p_{j_1,2}p_{j_2,1}}{p_{j_1,0}p_{j_2,1} + p_{j_1,1}p_{j_2,0} + p_{j_1,2}p_{j_2,2}} \end{aligned}$$

depends only on the first two coordinates of the vector $\underline{\mathbf{j}} = (j_1, j_2, j_3, j_4)$, which is a permutation of the set $\{1, 2, 3, 4\}$. It is easily seen that $R(1,3) = 46/27 > 47/34 = R(1,4)$ and $R(2,3) = 2 < 49/22 = R(2,4)$. Thus, whether $R(j_1, 3)$ is less, or greater than $R(j_1, 4)$, depends on the value of j_1. In other words, the best value of j_2 depends on that chosen for j_1. □

It would be interesting to prove or to find a counter example for (7) under the constraint (8), where

$$p_i = \max\{p_{i,j} \mid 1 \leq j < q\} < 1/q, \qquad 1 \leq i \leq n.$$

If (7) were true under these assumptions, then it might be possible to analyze Algorithm P1 (see Section 2) in the non–binary case.

2 An analysis of two probabilistic decoding algorithms for binary linear codes

In this section a method is given for the analysis of the probabilistic decoding algorithm for linear codes described in the Introduction (Algorithm P1, see below). This method is based on a sufficient condition of convergence of an iteratively computed sequence of error–vector probability distributions (Theorem 2). The proof of Theorem 2 is carried out using a special case of Theorem 1 for $q = 2$, which is a generalization of Sullivan inequality [11]. This approach enables one to estimate experimentally the dependence of the successful decoding probability on the DMC noise level, for an arbitrary binary code and for an arbitrary family of parity–check sets (the family which allows effective computation of *a posteriori* error probabilities). The method is illustrated by two examples. Another probabilistic decoding algorithm, applicable to "low–density parity–check codes" [6], is also analyzed.

Let C be a binary linear (n, k) code with a parity–check matrix $\underline{\mathbf{H}}$. The effect of a DMC with error probabilities $p_1, p_2, \ldots, p_n$ can be modeled by an n–dimensional binary random variable $\underline{\mathbf{E}}$ defined over $V_n = \{0, 1\}^n$, with independent coordinates and with probability distribution

$$P_{\underline{\mathbf{p}}}\{\underline{\mathbf{E}} = \underline{\mathbf{e}}\} = \prod_{i=1}^{n} p_i^{e_i}(1-p_i)^{1-e_i}, \qquad \underline{\mathbf{e}} \in V_n. \tag{9}$$

The real vector $\underline{\mathbf{p}} = (p_1, p_2, \ldots, p_n) \in [0,1]^n$ with the coordinates $p_i = P_{\underline{\mathbf{p}}}\{E_i = 1\}$, $1 \leq i \leq n$, determining the probability distribution of $\underline{\mathbf{E}}$, will be referred to as the error probability vector of the DMC (or: the error probability vector of the random variable $\underline{\mathbf{E}}$). Applying the codeword $\underline{\mathbf{x}} \in C$ to the input of the DMC, we get the random variable $\underline{\mathbf{Y}} = \underline{\mathbf{E}} + \underline{\mathbf{x}}$, where addition is operation in $\mathrm{GF}(2^n)$.

We shall now describe more precisely the first of the two probabilistic decoding algorithms to be discussed, which will be referred to as Algorithm P1. Suppose that for each codeword coordinate we have chosen a set of parity checks from the dual code. The vectors corresponding to the chosen parity checks have to be linearly independent in every set, see for example [2]. For all i, $1 \leq i \leq n$, define $\underline{\mathbf{H}}^{(i)}$ as the matrix whose rows are equal to the dual code codewords, corresponding to the chosen i–th set of parity checks. Let $\underline{\mathbf{y}}$, the *received message*, be a realization of the random variable $\underline{\mathbf{Y}}$. The function $\mathcal{F}_{\underline{\mathbf{y}}} : [0,1]^n \rightarrow [0,1]^n$ is defined as the mapping transforming the vector $\underline{\mathbf{p}}$ of the *a priori* error probabilities into the vector $\underline{\mathbf{P}}$ of *a posteriori* error probabilities, i.e.

$$\begin{aligned} P_i &= P_{\underline{\mathbf{p}}}\left(\{E_i = 1\} \mid \{\underline{\mathbf{H}}^{(i)}\underline{\mathbf{E}} = \underline{\mathbf{H}}^{(i)}\underline{\mathbf{y}}\}\right) \\ &= \frac{P_{\underline{\mathbf{p}}}\{\underline{\mathbf{H}}^{(i)}\underline{\mathbf{E}} = \underline{\mathbf{H}}^{(i)}\underline{\mathbf{y}}, E_i = 1\}}{P_{\underline{\mathbf{p}}}\{\underline{\mathbf{H}}^{(i)}\underline{\mathbf{E}} = \underline{\mathbf{H}}^{(i)}\underline{\mathbf{y}}\}}\ , \qquad 1 \leq i \leq n. \end{aligned} \tag{10}$$

Define the vector sequence $\{\underline{\mathbf{P}}^{(j)}\}_{j \geq 0}$ by the DMC error probability vector $\underline{\mathbf{p}} = \underline{\mathbf{P}}^{(0)}$, and by the recurrent relation

$$\underline{\mathbf{P}}^{(j+1)} = \mathcal{F}_{\underline{\mathbf{y}}}(\underline{\mathbf{P}}^{(j)}),\ j \geq 0. \tag{11}$$

To decode a received message by Algorithm P1 means to estimate the error vector by the vector $\underline{\bar{\mathbf{e}}}$, obtained from $\underline{\mathbf{P}}^{(d)}$ (for some fixed integer d) by rounding its coordinates to one binary digit, i.e.

$$\bar{e}_i = \begin{cases} 0, \ P_i^{(d)} \leq 1/2 \\ 1, \ P_i^{(d)} > 1/2 \end{cases} . \tag{12}$$

If the vector $\underline{\bar{\mathbf{e}}} + \underline{\mathbf{y}}$ is equal to a codeword $\underline{\mathbf{x}}$ that could have been applied to the DMC input, then decoding is successful, otherwise it is not.

If $\underline{\mathbf{p}}$ incorporates the reliability information related to the received message $\underline{\mathbf{y}}$, obtained by hard–decision, then Algorithm P1 is in fact a soft–decision decoding algorithm. A more sophisticated version of Algorithm P1 includes some information set decoding algorithm [4, pp. 102–131] as a second phase, which means that the result of decoding is the codeword $\underline{\bar{\mathbf{x}}}$ agreeing with $\underline{\bar{\mathbf{e}}}+\underline{\mathbf{y}}$ on the chosen information set. In this paper only the basic version of the algorithm is considered.

For $d = 1$ Algorithm P1 is in fact a known symbol–by–symbol decoding algorithm, see for example [7]. Repeated calculation of the *a posteriori* error probabilities is heuristically motivated, and enables to incorporate information from a large part of the received message into the decision on every error bit, see [6]. It is known that probabilistic decoding methods have low numerical complexity when orthogonal parity–check sets are used (parity checks with exactly one common member).

If the vector sequence $\{\underline{\mathbf{P}}^{(j)}\}_{j \geq 0}$ converges, then for large enough $d_0 > 0$ the vector $\underline{\bar{\mathbf{e}}}$, obtained by rounding the coordinates of $\underline{\mathbf{P}}^{(d)}$ according to (12), does not depend on d for $d > d_0$. The following theorem gives a sufficient condition for convergence of this vector sequence.

Theorem 2 *Suppose C is a linear (n, k) code with parity–check matrix $\underline{\mathbf{H}}$.*

For $1 \le i \le n$ *let the rows of matrix* $\underline{\mathbf{H}}^{(i)}$ *be some linearly independent vectors from the dual code. Let* $\underline{\mathbf{y}} \in V_n$ *be the received message, and let the vector sequence* $\{\underline{\mathbf{P}}^{(j)}\}_{j\ge 0}$ *be defined by the DMC error probability vector* $\underline{\mathbf{P}}^{(0)}$*, and by the recurrent relation (11). If for some* $\underline{\mathbf{x}} \in C$ *and for some* $d > 0$ *the coordinates of* $\underline{\mathbf{P}}^{(d)}$ *satisfy the condition*

$$P_i^{(d)} \begin{cases} < 1/2, & y_i = x_i \\ > 1/2, & y_i \ne x_i \end{cases}, \qquad 1 \le i \le n, \tag{13}$$

then

$$\lim_{j\to\infty} P_i^{(j)} \begin{cases} = 0, & y_i = x_i \\ = 1, & y_i \ne x_i \end{cases}, \qquad 1 \le i \le n. \qquad \square$$

The idea of the proof (see the Appendix) is to reduce the problem by an appropriate substitution to the case when the received message equals the all–zero vector. Then, using Theorem 1, it is proved that the coordinates of the transformed error probability vectors uniformly tend to zero (at least exponentially).

Suppose that $\underline{\mathbf{x}}$ in the statement of Theorem 2 is the transmitted codeword (this is often the case for some $d > 0$ when the noise level is low). If conditions of the theorem are satisfied then the limit of $\{\underline{\mathbf{P}}^{(j)}\}_{j\ge 0}$ is the error vector $\underline{\mathbf{e}} = \underline{\mathbf{x}} + \underline{\mathbf{y}}$. The converse claim obviously holds: if $\lim_{j\to+\infty} \underline{\mathbf{P}}^{(j)} = \underline{\mathbf{e}}$, then there exists $d > 0$, such that (13) is satisfied. Therefore, decoding by Algorithm P1 is successful for large enough d if, and only if, $\lim_{j\to+\infty} \underline{\mathbf{P}}^{(j)} = \underline{\mathbf{e}}$.

During decoding by Algorithm P1 computation of vectors from the sequence $\{\underline{\mathbf{P}}^{(s)}\}_{1\le s\le d}$ can be interrupted for some $s < d$ if (13) is satisfied with d replaced by s.

Let us now discuss the significance of Theorem 2. For $d = 1$ Theorem 2 can be stated as follows: if it is possible to decode a received message $\underline{\mathbf{y}}$

on a symbol–by–symbol basis using the parity–check matrices $\underline{\mathbf{H}}^{(i)}$, $1 \leq i \leq n$, then applying Algorithm P1 to the same message $\underline{\mathbf{y}}$ also leads to successful decoding. The importance of Theorem 2 arises from the fact that Algorithm P1 leads to successful decoding also if (13) is satisfied for an arbitrary $d \geq 1$.

Exact calculation of the successful decoding probability (using Algorithm P1) is practically impossible, except for simple, useless codes. But it is possible to estimate this probability experimentally. Let us first consider the case of a BSC, where $p_1 = p_2 = \cdots = p_n = p < 1/2$. Let $\underline{\mathbf{x}}$ denote the codeword applied to the input of the BSC. For $d \geq 1$ let A_d denote the set of error vectors $\underline{\mathbf{e}} \in V_n$ with the following property: if Algorithm P1 is applied to $\underline{\mathbf{y}} = \underline{\mathbf{x}} + \underline{\mathbf{e}}$, then $\underline{\mathbf{P}}^{(d)}$ satisfies (13). From the proof of Theorem 2 we have $A_1 \subseteq A_2 \subseteq \cdots$. Define the random variable U_d as the indicator of the event $\underline{\mathbf{E}} \in A_d$, i.e.

$$U_d = \begin{cases} 1, & \underline{\mathbf{E}} \in A_d \\ 0, & \underline{\mathbf{E}} \notin A_d \end{cases}, \qquad d \geq 1. \tag{14}$$

The quantity

$$\alpha_d = P\left\{U_d = 1\right\} = P\{\underline{\mathbf{E}}_d \in A_d\}, \qquad d \geq 1, \tag{15}$$

is the probability that $\underline{\mathbf{P}}^{(d)}$ satisfies (13), i.e. the probability of successful decoding using $\underline{\mathbf{P}}^{(d)}$. The probability α_d can be estimated statistically in the usual way. If the independent random variables $\underline{\mathbf{E}}^{(s)}$, $1 \leq s \leq N$, $N > 1$, are distributed as $\underline{\mathbf{E}}$ in (9), then the corresponding random variables $U_d^{(s)}$ have the same probability distribution. The random variable $N\bar{U}_d$, where $\bar{U}_d = \frac{1}{N}\sum_{s=1}^{N} U_d^{(s)}$, is binomially distributed with parameters α_d and N. Therefore, the probability α_d can be estimated as $\alpha_d \simeq \bar{u}_d$, where $\bar{u}_d$ is the realization of the random variable $\bar{U}_d$. More precisely, the quantity α_d lies in the interval $[\bar{u}_d - 2\sigma_d, \bar{u}_d + 2\sigma_d]$, with a probability of approximately 95%, where $\sigma_d \simeq \sqrt{\bar{u}_d(1-\bar{u}_d)/N}$. Note that the outcome of the experiment does not depend on the choice of the codeword $\underline{\mathbf{x}}$, and so without loss of generality it can be taken $\underline{\mathbf{x}} = \underline{\mathbf{0}}$. The described method is illustrated by the following example.

Example 2 Let C be a linear $(512, 100)$ code (the parameters n and k have similar values as in an example in [6]), whose codewords are vectors $\underline{\mathbf{x}} \in V_{512}$ satisfying the following parity checks

$$x_i + x_{i+37} + x_{i+100} = 0, \quad 1 \le i \le 412. \tag{16}$$

This is a recurrence relation with the characteristic polynomial $f(z) = 1 + z^{37} + z^{100}$. The sequence $x_1, x_2, \ldots$ also satisfies recurrence relation with the characteristic polynomial $f(z)g(z)$ for an arbitrary binary polynomial $g(z)$. In particular, for $g(z) = f(z)$ and $g(z) = (f(z))^3$, we get characteristic polynomials $(f(z))^2 = f(z^2)$ and $(f(z))^4 = f(z^4)$, and so codewords of C also satisfy the parity checks

$$x_i + x_{i+74} + x_{i+200} = 0, \quad 1 \le i \le 312, \tag{17}$$

and

$$x_i + x_{i+148} + x_{i+400} = 0, \quad 1 \le i \le 112. \tag{18}$$

Codewords of C are easily produced using an appropriate linear feedback shift register. For any i, $1 \le i \le 512$, the parity–check matrix $\underline{\mathbf{H}}^{(i)}$ consists of all parity checks of the form (16), (17) and (18), containing the coordinate x_i of the codeword (this construction is used in [8]). It can easily be verified that the parity–check sets corresponding to these matrices are linearly independent and orthogonal (i.e. that all columns of the matrix $\underline{\mathbf{H}}^{(i)}$, excluding the i–th, contain exactly one 1).

Suppose that we fix a codeword $\underline{\mathbf{x}} \in C$, the BSC transition probability p and the number d of iterations in Algorithm P1. The realizations $\underline{\mathbf{e}}^{(s)} \in V_{512}$, of random variables $\underline{\mathbf{E}}^{(s)},\ 1 \leq s \leq N = 100$, can be obtained using a random number generator. Algorithm P1 is applied to every received message $\underline{\mathbf{y}}^{(s)} = \underline{\mathbf{x}} + \underline{\mathbf{e}}^{(s)},\ 1 \leq s \leq N$, and the number of successful decodings is counted. The dependence of the estimated probability of successful decoding on the error probability p is depicted in Figure 1 for $p = 8/256(1/256)50/256$ and $d = 1, 2, 3, 5, 7, 10$. One can see that the probability of successful decoding is close to one for $p < 1/16 = 0.0625$ and that this probability decreases rapidly for larger values of p. $\square$

Consider now a DMC related as follows to a channel with binary antipodal signals and white Gaussian noise, with optimum demodulation (see [2] for example). For the input bit $x \in \{0, 1\}$, denote by Ξ the random variable equal to the output of demodulator. Let $(-1)^x S$ be its output in the absence of noise, and let σ^2 be its variance when noise is present. The probability that Ξ lies in the interval $(\xi, \xi + d\xi)$ is

$$(2\pi\sigma^2)^{-1/2} \exp\left(-(\xi - (-1)^x S)^2/2\sigma^2\right) d\xi. \tag{19}$$

If ξ is the demodulator output, then the received bit y is obtained by hard decision

$$y = \begin{cases} 0, & \xi \geq 0 \\ 1, & \xi < 0 \end{cases}, \tag{20}$$

and the probability of erroneous decision is

$$\left(1 + \exp(4\gamma|\xi|/S)\right)^{-1}. \tag{21}$$

Here $\gamma = S^2/(2\sigma^2)$ denotes the signal–to–noise ratio (SNR). Repeating this procedure for each codeword bit x_i and the demodulator output ξ_i, $1 \leq i \leq n$, we get the received message $\underline{\mathbf{y}}$ and the DMC error probability vector $\underline{\mathbf{p}}$. The average error probability for such a DMC is $\bar{p} = Q(S/\sigma) = Q(\sqrt{2\gamma})$, where

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} \exp(-z^2/2)\, dz.$$

As in the case of a BSC, the probability of successful decoding can be estimated statistically. Here the elementary events space is $[0,1]^n \times V_n$, because $\underline{\mathbf{p}} = \underline{\mathbf{P}}^{(0)}$ is a realization of a random vector, determining the conditional probability distribution of the error vector $\underline{\mathbf{E}}$. For fixed $d \geq 1$ denote by A_d the set of pairs $(\underline{\mathbf{P}}^{(0)}, \underline{\mathbf{e}})$ with the following property: if Algorithm P1 is applied to $\underline{\mathbf{y}} = \underline{\mathbf{x}} + \underline{\mathbf{e}}$, then $\underline{\mathbf{P}}^{(d)}$ satisfies (13). Like before we can define random variable U_d (14) and the probability α_d (15). This probability is then estimated by the average of the realizations of $N \geq 1$ independent random variables, with the same probability distribution as that of U_d.

Example 3 Let C be the linear $(512, 100)$ code of Example 2. For fixed $\underline{\mathbf{x}} \in C$, γ and d we generate random normal deviates ξ_i, $1 \leq i \leq n$, according to distribution (19), where $x = x_i$ and $S = 1$. These deviates can be treated as outputs from a Gaussian channel, when $\underline{\mathbf{x}}$ is applied to its input. By hard decision (20) we get the received message $\underline{\mathbf{y}}$ (the output of the related DMC). The DMC error probability vector $\underline{\mathbf{p}} = \underline{\mathbf{P}}^{(0)}$ is obtained using (21) with ξ_i substituted for ξ, $1 \leq i \leq n$. The realization of the error vector is obviously $\underline{\mathbf{e}} = \underline{\mathbf{x}} + \underline{\mathbf{y}}$. Algorithm P1 is then applied to $\underline{\mathbf{y}}$ and $\underline{\mathbf{P}}^{(0)}$. Repeating this procedure $N = 100$ times, we estimate the probability of successful decoding by the quotient of the number of successful decodings and N. The results of such an experiment for $10 \log \gamma = -8(0.2)4\,\text{dB}$ and $d = 1, 2, 3, 5, 7, 10$ are displayed in Figure 2. It is seen that the probability of successful decoding is close to one if $10 \log \gamma \geq -6\,\text{dB}$ (for $d = 10$). The average error probability corresponding to this SNR is $\bar{p} = Q(S/\sigma) = Q(\sqrt{2\gamma}) = 15.8\%$. Note that the results for BSC (Figure 1) are similar to those in Figure 2, if SNR is replaced by $\gamma' = 2\gamma$ and then transformed into the average error probability $\bar{p} = Q(\sqrt{2\gamma'})$. This fact arises from the use of reliability information (contained in the demodulator output) in the second case. □

The coordinates of $\underline{\mathbf{P}}^{(1)}$ in Algorithm P1 have an obvious interpretation, because they are the *a posteriori* error probabilities. But the coordinates of other vectors $\underline{\mathbf{P}}^{(2)}, \ldots$ may not be considered as the actual *a posteriori* probabilities inasmuch as they fail to take into account the dependence between the error terms which results from the previous decoding step. However, they are dealt with in the algorithm as such probabilities, which is both heuristically justified and practically successful. Still, this is in fact a "theoretical" drawback of Algorithm P1. The probabilistic algorithm for decoding "low–density parity–check codes" [6] (which will be referred to as Algorithm P2) is more precise in this sense. Algorithm P2 is outlined here somewhat more formally, and then it is analyzed similarly to Algorithm P1.

A code C is a binary (n, j, k) *low–density parity–check code* if every row of its parity–check matrix $\underline{\mathbf{H}}$ has exactly k ones, and if every column of $\underline{\mathbf{H}}$ has exactly j ones. Let R be the number of rows of $\underline{\mathbf{H}}$ and let

$$\mathcal{H}_u = \{(r,i) \mid 1 \leq i \leq n,\ 1 \leq r \leq R,\ H_{r,i} = u\} \qquad u \in \{0,1\}. \qquad (22)$$

If $(r,i) \in \mathcal{H}_1$ then let $\underline{\mathbf{H}}^{(i,r)}$ ($\underline{\mathbf{H}}^{(i)}$) denote the $(j-1) \times n$ $\quad (j \times n)$ matrix, consisting of the rows of $\underline{\mathbf{H}}$ with the index ρ satisfying $H_{\rho,i} = 1$ and $\rho \neq r$ ($H_{\rho,i} = 1$). We assume that the parity checks in $\underline{\mathbf{H}}^{(i)}$ for the i–th bit are orthogonal, $1 \leq i \leq n$.

Suppose $\underline{\mathbf{y}}$ is an arbitrary received message. Let $\underline{\mathbf{p}}$ be a DMC error probability vector and let $\underline{\mathbf{b}}$ be an $R \times n$ matrix with elements in $[0,1]$, where $b_{r,i} = 0$ for all $(r,i) \in \mathcal{H}_0$. For the fixed $(r,i) \in \mathcal{H}_1$ the vector $\underline{\mathbf{g}} = \underline{\mathbf{g}}(\underline{\mathbf{H}}, \underline{\mathbf{b}}, r, i)$ is defined by

$$g_\iota = \begin{cases} p_i, & \iota = i \\ b_{\rho,\iota}, & (\rho,\iota) \in \mathcal{H}_1,\ H_{\rho,i} = 1,\ \rho \neq r,\ \iota \neq i \\ 0, & \text{in other cases} \end{cases} . \tag{23}$$

The coordinates of $\underline{\mathbf{g}}$ corresponding to the non–zero columns of $\underline{\mathbf{H}}^{(i,r)}$ are uniquely determined here, because the parity checks in $\underline{\mathbf{H}}^{(i,r)}$ of the bit i are orthogonal. Let $B_{r,i}$ denote the conditional probability

$$B_{r,i} = P_{\underline{\mathbf{g}}}\left(\{E_i = 1\} \mid \{\underline{\mathbf{H}}^{(i,r)}\underline{\mathbf{E}} = \underline{\mathbf{H}}^{(i,r)}\underline{\mathbf{y}}\}\right) \tag{24}$$

Note that $B_{r,i}$ depends only on the coordinates of $\underline{\mathbf{g}}$ corresponding to the non–zero columns of $\underline{\mathbf{H}}^{(i,r)}$. If $(r,i) \in \mathcal{H}_0$ then let $B_{r,i} = 0$. Denote by $\Phi_{\underline{\mathbf{y}},\underline{\mathbf{p}}}$ the mapping transforming $\underline{\mathbf{b}}$ into $\underline{\mathbf{B}} = [B_{r,i}]_{R \times n}$. Function Φ here plays the same role as function $\mathcal{F}$ given by (10) in Algorithm P1.

Define the sequence of $R \times n$ matrices $\{\underline{\mathbf{B}}^{(j)}\}_{j \geq 0}$ by $\underline{\mathbf{B}}^{(0)}$, the matrix with the elements $B_{r,i}^{(0)} = H_{r,i} p_i$, and by the recurrent relation

$$\underline{\mathbf{B}}^{(s+1)} = \Phi_{\underline{\mathbf{y}},\underline{\mathbf{p}}}(\underline{\mathbf{B}}^{(s)}),\ s \geq 0. \tag{25}$$

For $(r,i) \in \mathcal{H}_0$ and $s \geq 0$ we obviously have $B_{r,i}^{(s)} = 0$. In Algorithm P2 the matrix $\underline{\mathbf{B}}^{(d)}$ is used to calculate the vector of *a posteriori* error probabilities $\underline{\tilde{\mathbf{P}}}^{(d)}$ for some $d > 0$, where

$$\tilde{P}_i^{(d)} = P_{\underline{\mathbf{g}}'}\left(\{E_i = 1\} \mid \{\underline{\mathbf{H}}^{(i)}\underline{\mathbf{E}} = \underline{\mathbf{H}}^{(i)}\underline{\mathbf{y}}\}\right).$$

Here the probability distribution vector $\underline{\mathbf{g}}' = \underline{\mathbf{g}}(\underline{\mathbf{H}}, \underline{\mathbf{b}}, i)$ is defined (see (23)) by

$$g'_\iota = \begin{cases} p_i, & \iota = i \\ B^{(d)}_{\rho,\iota}, & (\rho,\iota) \in \mathcal{H}_1,\ H_{\rho,i} = 1,\ \iota \neq i \\ 0, & \text{in other cases} \end{cases} .$$

In the general case the quantities $\tilde{P}^{(d)}_i$ are not the actual *a posteriori* error probabilities. Still, when $\underline{\mathbf{H}}$ is a parity–check matrix of a low–density parity–check code and d is small enough, $\tilde{P}^{(d)}_i$ *are* the *a posteriori* error probabilities, see [5]. The last step of Algorithm P2 is to calculate the error vector estimate $\underline{\bar{\mathbf{e}}}$ using $\underline{\tilde{\mathbf{P}}}^{(d)}$, like in Algorithm P1

$$\bar{e}_i = \begin{cases} 0, & \tilde{P}^{(d)}_i \le 1/2 \\ 1, & \tilde{P}^{(d)}_i > 1/2 \end{cases}, \qquad 1 \le i \le n. \tag{26}$$

If the vector $\underline{\bar{\mathbf{e}}} + \underline{\mathbf{y}}$ is equal to a codeword $\underline{\mathbf{x}}$ that could have been applied to the DMC input, then decoding is successful, otherwise it is not.

Define a continuous increasing function $f_{j,k,u} : [0, 1/2] \longrightarrow [0, u]$ by

$$f_{j,k,u}(t) = \begin{cases} 0, & t = 0 \\ \left(1 + \frac{1-u}{u}\left(\frac{1+(1-2t)^{k-1}}{1-(1-2t)^{k-1}}\right)^{j-1}\right)^{-1}, & 0 < t \le 1/2 \end{cases},$$

where $j, k \ge 3$, and $0 < u < 1$. For $t \longrightarrow 0^+$ we have $f_{j,k,u}(t) = O(t^{j-1})$, and therefore the inequality $f_{j,k,u}(t) < t$ holds for small enough $t > 0$. Denote by $t_0 = t_0(j,k,u)$ the lowest upper bound on the values of t such that $f_{j,k,u}(t') < t'$ for all $t' < t$. Obviously, $t_0(j,k,u)$ is equal to $1/2$, or it is equal to the smallest positive zero of the function $f_{j,k,u}(t) - t$. In both cases we have the inequality

$$f_{j,k,u}(t) < t, \qquad 0 < t < t_0(j,k,u). \tag{27}$$

The value of $t_0(j,k,u)$ is a non–increasing function of u. Some values of $t_0(j,k,u)$ are listed in the following table.

(j,k)	$(3,4)$	$(3,5)$	$(3,6)$	$(4,6)$
u				
0.90	0.012205	0.006900	0.004426	0.029258
0.95	0.005815	0.003279	0.002101	0.020208
0.99	0.001121	0.000631	0.000404	0.008918

We are now ready to formulate a sufficient condition for convergence of the matrix sequence $\{\underline{\mathbf{B}}^{(s)}\}_{s\geq 0}$.

Theorem 3 *Let C be an (n,j,k) low–density parity–check code. Let $\underline{\mathbf{H}}^{(i,r)}$ denote the parity–check matrices used in Algorithm P2 and let the matrix sequence $\{\underline{\mathbf{B}}^{(s)}\}_{s\geq 0}$ be defined by (25), where $\underline{\mathbf{B}}^{(0)}$ is the matrix with the elements $B^{(0)}_{r,i} = H_{r,i}p_i$, $1 \leq i \leq R$, $1 \leq i \leq n$. Denote*

$$p = \min\{p_j \mid 1 \leq j \leq n\},$$

and let $t_0 = t_0(j,k,1-p)$. If $\underline{\mathbf{y}} \in V_n$ is the received message and if for some $\underline{\mathbf{x}} \in C$, $d \geq 0$ for all $(r,i) \in \mathcal{H}_1$ (22) we have

$$B^{(d)}_{r,i} \begin{cases} < t_0, & y_i = x_i \\ > 1 - t_0, & y_i \neq x_i \end{cases}, \tag{28}$$

then

$$\lim_{s\to\infty} B^{(s)}_{r,i} \begin{cases} = 0, & y_i = x_i \\ = 1, & y_i \neq x_i \end{cases}, \qquad (r,i) \in \mathcal{H}_1. \qquad \square \tag{29}$$

The proof is similar to that of Theorem 2 (see the Appendix).

Suppose $\underline{\mathbf{x}}$ is the transmitted codeword. Then, as in Theorem 2, we have $\lim_{s\to+\infty} B^{(s)}_{r,i} = e_i = x_i + y_i$ for $1 \leq i \leq n$ if, and only if, for some $d > 0$ condition (28) is satisfied. If the assumptions of Theorem 3 are satisfied, then there exists d_0 (not necessary equal to d), such that for all $d' > d_0$ the vector $\bar{\underline{\mathbf{e}}}$ from (26) (with d replaced by d') is equal to $\underline{\mathbf{e}}$, i.e. decoding by Algorithm P2 is successful. One can estimate statistically the probability that the conditions of Theorem 3 are satisfied (the probability which is a lower bound on the successful decoding probability when d is large enough) similarly to the case of Algorithm P1.

Acknowledgement

The author wishes to express his gratitude to professor Ž. Mijajlović, who encouraged him to investigate the convergence of repeatedly computed *a posteriori* error probabilities. The author is grateful to anonymous referees for many useful remarks, and also for the suggestion to generalize Theorems 2 and 3 to the case of DMC (BSC with different error probabilities).

Appendix

Proof of Theorem 1

In the proof of Theorem 1 the following lemma, an immediate generalization of Massey lemma in [11] (where the case $q = 2$ is treated), is used. The element $\underline{\mathbf{v}}$ from a coset K of subgroup G is a coset leader of K if it does not contain any element of lower weight.

Lemma 1 *Let G be an arbitrary subgroup of the additive group of the field* $\mathrm{GF}(q^n)$*, (q a power of a prime), and let K be a proper coset of G. Suppose that the element $\underline{\mathbf{v}}$ is a coset leader of K. Denote by G' the subgroup obtained* *from G by replacing the coordinates of its elements, corresponding to non-zero coordinates of $\underline{\mathbf{v}}$, by zeros. Then the order G' is equal to the order of G.* □

Proof. Suppose that Lemma 1 is not true, i.e. that there exist elements $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in G$, $\underline{\mathbf{a}} \neq \underline{\mathbf{b}}$, such that replacing the coordinates corresponding to non-zero coordinates of $\underline{\mathbf{v}}$ by zero, we get the same element $\underline{\mathbf{c}} \in G'$. Then we have $\underline{\mathbf{a}} - \underline{\mathbf{b}} \in G$, and all the coordinates of $\underline{\mathbf{a}} - \underline{\mathbf{b}}$, corresponding to the zero coordinates of $\underline{\mathbf{v}}$, are equal to zero. Let β be an arbitrary non–zero coordinate of $\underline{\mathbf{a}} - \underline{\mathbf{b}}$, and let γ be the corresponding (non–zero) coordinate of $\underline{\mathbf{v}}$. Then the weight of the coset element $-\gamma\beta^{-1}(\underline{\mathbf{a}} - \underline{\mathbf{b}}) + \underline{\mathbf{v}}$ is smaller than that of $\underline{\mathbf{v}}$, which contradicts the assumption. So, Lemma 1 is proved. □

The subgroup G of the additive group of $\mathrm{GF}(q^n)$ is a linear subspace of the linear space V_n, i.e. a linear code. Let $\underline{\mathbf{H}}$ be a parity–check matrix of the code G, (matrix whose rows form a basis of its dual code). The set of indices $(i_1, i_2, \ldots, i_k)$ is an *information set* of the linear code G if the columns of $\underline{\mathbf{H}}$ with indices from this set are linearly independent. All the coordinates of a codeword can be expressed as linear combinations of coordinates with indices from the information set. Lemma 1 has the following obvious corollary.

Corollary 1 *If vector $\underline{\mathbf{v}}$ is a coset leader of a coset K of a linear code G, then there exists an information set of G which is a subset of the set of indices of the zeros in $\underline{\mathbf{v}}$.* □

We are now ready to start with the proof of Theorem 1. The statement of the theorem will be proved by induction over k, $0 \leq k < n$, where q^k is the order of the subgroup G. For $k = 0$ we have $G = \{\underline{\mathbf{0}}\}$ and $K = \{\underline{\mathbf{v}}\}$, where $\underline{\mathbf{0}}$ denotes the vector whose all coordinates are equal to zero. Without loss of generality suppose that the coordinates of elements of the linear space V_n are permuted so that $v_i = 0$ for $1 \leq i \leq n-m$ and $v_i \neq 0$ for $i > n-m$, where $m = w(\underline{\mathbf{v}})$, the Hamming weight of $\underline{\mathbf{v}}$. Then we have

$$\begin{aligned} \frac{P(G)}{P(K)} &= \frac{\prod_{i=1}^{n}(1-(q-1)p_i)}{\prod_{i=1}^{n-m}(1-(q-1)p_i)\prod_{i=n-m+1}^{n}p_i} \\ &= \prod_{i=n-m+1}^{n}\frac{1-(q-1)p_i}{p_i} > \frac{1-(q-1)p_n}{p_n} \\ &> \frac{1-(q-1)p_{l_1}}{p_{l_1}} = F_0(\underline{\mathbf{p}}). \end{aligned}$$

The permutation of indices $\underline{\mathbf{l}}$ is defined by the ordering (6).

Suppose that the statement of Theorem 1 is proved for all subgroups of order not greater than q^{k-1}. The inequality (7) for subgroups of order q^k and their cosets will be proved by induction over the coset leader weight m. Let G be an arbitrary subgroup of order q^k, and let K be an arbitrary coset of it, with the element $\underline{\mathbf{v}}$ as one of the coset leaders, $w(\underline{\mathbf{v}}) = m$.

Consider the case $m = 1$ first. Without loss of generality suppose that the set $\{1, 2, \ldots, k\}$ is an information set of the code, and that v_{k+1} is the only non–zero coordinate of $\underline{\mathbf{v}}$ (this can be achieved by an appropriate permutation of coordinates). Let $\underline{\mathbf{x}}' = [x_1\ x_2\ \ldots\ x_k]^T$. Denote

$$\underline{\mathbf{g}}(\underline{\mathbf{x}}') = [x_1\ x_2\ \ldots\ x_k\ L_{k+1}\ \ldots\ L_n]^T\,, \tag{30}$$

and

$$\underline{\mathbf{c}}(\underline{\mathbf{x}}') = \underline{\mathbf{g}}(\underline{\mathbf{x}}') + \underline{\mathbf{v}} = [x_1\ x_2\ \ldots\ x_k\ L_{k+1} + v_{k+1}\ \ldots\ L_n]^T\,. \tag{31}$$

Here $L_i = L_i(\underline{\mathbf{x}}')$, $k+1 \leq i \leq n$, are some linear combinations of the independent variables $x_1, x_2, \ldots, x_k$. Obviously, we can write

$$G = \{\underline{\mathbf{g}}(\underline{\mathbf{x}}') \mid \underline{\mathbf{x}}' \in V_k\}, \tag{32}$$

and

$$K = \{\underline{\mathbf{c}}(\underline{\mathbf{x}}') \mid \underline{\mathbf{x}}' \in V_k\}. \tag{33}$$

The set V_k is partitioned in q disjoint subsets $\mathcal{D}_u,\ u \in V_1$, according to the value of the linear combination $L_{k+1}(\underline{\mathbf{x}}'),\ \underline{\mathbf{x}}' \in V_k$,

$$\mathcal{D}_u = \{\underline{\mathbf{x}}' \in V_k \mid L_{k+1}(\underline{\mathbf{x}}') = u\}, \quad u \in V_1.$$

Denote

$$T_u = \sum_{\underline{\mathbf{x}}' \in \mathcal{D}_u} \left(\prod_{i=1}^{k} p_{i,x_i} \right) \left(\prod_{i=k+2}^{n} p_{i,L_i} \right), \qquad u \in V_1,$$

where $p_{i,j},\ 1 \le i \le n,\ j \in V_1$, are given by (2). The probabilities $p_{i,0}$ are denoted by $q_i,\ 1 \le i \le n$.

The set $G_0 = \{\underline{\mathbf{g}}(\underline{\mathbf{x}}') \mid \underline{\mathbf{x}}' \in \mathcal{D}_0\}$ is a subgroup of G. Suppose first that $G_0 = G$, i.e. $L_{k+1}(\underline{\mathbf{x}}') \equiv 0,\ \underline{\mathbf{x}}' \in V_k$. According to Property 2 of the function F we have

$$\frac{P(K)}{P(G)} = \frac{p_{k+1}}{q_{k+1}} \le \frac{p_{j_1}}{q_{j_1}} = \phi_0(\underline{\mathbf{p}}) \le \phi_k(\underline{\mathbf{p}}),$$

which means that the inequality (7) is true in this case. Here we denoted $1/F_k(\underline{\mathbf{p}})$ by $\phi_k(\underline{\mathbf{p}})$ for simplicity.

Suppose now that $L_{k+1}(\underline{\mathbf{x}}') \ne 0$ for some $\underline{\mathbf{x}}' \in V_k$. The order of G_0 is then q^{k-1}, and the sets $G_u = \{\underline{\mathbf{g}}(\underline{\mathbf{x}}') \mid \underline{\mathbf{x}}' \in \mathcal{D}_u\},\ u \in V_1,\ u \ne 0$, are the cosets of G_0. Denote by G'_u the set obtained from the set G_u by deleting the $(k+1)$–th coordinate from all of its elements, $u \in V_1$, and let $\underline{\mathbf{p}}' = (p_1, \ldots, p_k, p_{k+2}, \ldots, p_n)$. According to Lemma 1, the order of the subgroup G'_0 is q^{k-1}. Since the sets $G'_u,\ u \ne 0$, are the cosets of G'_0, by the inductive hypothesis we get

$$T_u/T_0 \le \phi_{k-1}(\underline{\mathbf{p}}'), \qquad u \in V_1, u \ne 0, \tag{34}$$

because $P(G'_u) = T_u$. The probabilities $P(G)$ and $P(K)$ can be expressed by

$$P(G) = q_{k+1}T_0 + p_{k+1}T$$

and

$$P(K) = q_{k+1}T_{-v} + p_{k+1}\left(T + T_0 - T_{-v}\right),$$

where $T = \sum_{u \in V_1} T_u \; - T_0$ and $v = v_{k+1}$. Thus, we have

$$\frac{P(K)}{P(G)} = 1 - (q_{k+1} - p_{k+1})\,\frac{1 - (T_{-v}/T_0)}{q_{k+1} + p_{k+1}(T/T_0)}\ .$$

The right–hand side of this equality increases with both T_{-v}/T_0 and T/T_0. Therefore, using the inequalities (34) we get

$$\begin{aligned} \frac{P(K)}{P(G)} &\leq 1 - (q_{k+1} - p_{k+1})\,\frac{1 - \phi_{k-1}(\underline{\mathbf{p}}')}{q_{k+1} + p_{k+1}(q-1)\phi_{k-1}(\underline{\mathbf{p}}')} \qquad (35) \\ &= \frac{1 - (1 - p_{k+1})(1 - \phi_{k-1}(\underline{\mathbf{p}}'))}{1 - p_{k+1}(q-1)(1 - \phi_{k-1}(\underline{\mathbf{p}}'))}\ . \end{aligned}$$

Suppose that $j_1, j_2, \ldots, j_k$ are the indices in vector $\underline{\mathbf{p}}$ of the k largest coordinates of vector $\underline{\mathbf{p}}'$. Then we have

$$\phi_{k-1}(\underline{\mathbf{p}}') = \frac{1 - \Pi}{1 + (q-1)\Pi}\ ,$$

where $\Pi = \prod_{i=1}^{k}(1 - qp_{j_i})$. Substituting this in the right–hand side of (35) we get

$$\frac{P(K)}{P(G)} \leq \frac{1 - (1 - qp_{k+1})\Pi}{1 + (1 - qp_{k+1})(q-1)\Pi}\ .$$

Denote the right–hand side of this inequality by A. From the definition of the vector $\underline{\mathbf{p}}'$ we obviously have $j_i \neq k+1$ for $1 \leq i \leq k$. If p_{k+1} is one of the $k+1$ largest coordinates of $\underline{\mathbf{p}}$, then $A = \phi_k(\underline{\mathbf{p}})$. Otherwise we have

$$p_{k+1} \leq p_{l_{k+1}}$$

(the permutation $\underline{\mathbf{l}}$ is defined by (6)), and since A is a non-decreasing function of p_{k+1}, we get $A \leq \phi_k(\underline{\mathbf{p}})$. Thus, if the weight of the coset leader $\underline{\mathbf{v}}$ is one, then in both cases inequality (7) holds.

Suppose now that (7) is proved for subgroups G of order not exceeding q^{k-1}, and also when the order of G is q^k, but the weight of a coset leader of K is less than $m,\ m \geq 1$. Suppose G is an arbitrary subgroup of order q^k and K is any coset of G. Let $\underline{\mathbf{v}}$ be a coset leader of K, $w(\underline{\mathbf{v}}) = m$. Similarly to the case $m = 1$, we can assume without loss of generality that the set $\{1, 2, \ldots, k\}$ is an information set of G, and that the indices of the non-zero coordinates of $\underline{\mathbf{v}}$ are $k+1, k+2, \ldots, k+m$. Any element of G can be expressed by (30), and the elements of K are given by

$$\begin{aligned} \underline{\mathbf{c}}(\underline{\mathbf{x}}') &= \underline{\mathbf{g}}(\underline{\mathbf{x}}') + \underline{\mathbf{v}} \\ &= [x_1\ \ldots\ x_k\ L_{k+1} + v_{k+1}\ \ldots\ L_{k+m} + v_{k+m}\ \ldots\ L_n]^T,\ \underline{\mathbf{x}}' \in V_k, \end{aligned}$$

i.e. the equalities (32) and (33) hold. The set V_k is partitioned in q disjoint subsets $\mathcal{D}_u,\ u \in V_1$, according to the value of $L_{k+m}(\underline{\mathbf{x}}')$

$$\mathcal{D}_u = \{\underline{\mathbf{x}}' \in V_k \mid L_{k+m}(\underline{\mathbf{x}}') = u\}, \qquad u \in V_1.$$

Denote by T_u and $R_u,\ u \in V_1$, the sums

$$T_u = \sum_{\underline{\mathbf{x}}' \in \mathcal{D}_u} \left(\prod_{i=1}^{k} p_{i,x_i}\right) \left(\prod_{i=k+1}^{k+m-1} p_{i,L_i}\right) \left(\prod_{i=k+m+1}^{n} p_{i,L_i}\right),$$

and

$$R_u = \sum_{\underline{\mathbf{x}}' \in \mathcal{D}_u} \left(\prod_{i=1}^{k} p_{i,x_i}\right) \left(\prod_{i=k+1}^{k+m-1} p_{i,L_i+v_i}\right) \left(\prod_{i=k+m+1}^{n} p_{i,L_i}\right)$$

respectively, see (2). Then we have

$$P(G) = q_{k+m} T_0 + p_{k+m} T \tag{36}$$

and

$$P(K) = q_{k+m} R_{-v} + p_{k+m} \left(R - R_{-v} + R_0 \right), \tag{37}$$

where $T = \sum_{u \in V_1} T_u \; - T_0$, $R = \sum_{u \in V_1} R_u \; - R_0$ and $v = v_{k+m}$. Suppose G' is a subgroup obtained from G by replacing the $(k+m)$–th coordinate in every element of G by zero. Similarly, let K' be the set of vectors obtained from K by replacing the $(k+m)$–th coordinate in every element of K by zero. According to Lemma 1, the order of G' is q^k. The set K' is a coset of G', and the element

$$\underline{\mathbf{v}}' = [0 \;\ldots\; 0 \; v_{k+1} \;\ldots\; v_{k+m-1} \; 0 \;\ldots\; 0]^T$$

is one of its coset leaders (see for example [9, Theorem 3.9]). The weight of the vector $\underline{\mathbf{v}}'$ is $m-1$, and so by the inductive hypothesis we have

$$P(K')/P(G') \leq \phi_k(\underline{\mathbf{p}}') < 1, \tag{38}$$

where $\underline{\mathbf{p}}'$ is the vector obtained from $\underline{\mathbf{p}}$ by deleting its $(k+m)$–th coordinate. Substituting $P(G') = q_{k+m}(T_0 + T)$ and $P(K') = q_{k+m}(R_0 + R)$ in the preceding inequality, we obtain

$$\phi_k(\underline{\mathbf{p}})(T_0 + T) \geq R_0 + R, \tag{39}$$

because $\phi_k(\underline{\mathbf{p}}') \leq \phi_k(\underline{\mathbf{p}})$.

Consider the case $L_{k+m}(\underline{\mathbf{x}}') \equiv 0$, $\underline{\mathbf{x}}' \in V_k$, first. Then we have $\mathcal{D}_0 = V_k$, and consequently $G_0 = \{\underline{\mathbf{g}}(\underline{\mathbf{x}}') \mid \underline{\mathbf{x}}' \in \mathcal{D}_0\} = G$ and $T_u = R_u = 0$ for $u \neq 0$. Replacing $P(G)$ by $P(G')$ and $P(K)$ by $(p_{k+m}/q_{k+m})\, P(K')$, from (38) we get

$$\frac{P(K)}{P(G)} = \frac{p_{k+m}}{q_{k+m}} \frac{P(K')}{P(G')} < \frac{p_{k+m}}{q_{k+m}} \leq \frac{p_{j_1}}{q_{j_1}} = \phi_0(\underline{\mathbf{p}}) \leq \phi_k(\underline{\mathbf{p}}).$$

Suppose now that $L_{k+m}(\underline{\mathbf{x}}') \neq 0$ for some $\underline{\mathbf{x}}' \in V_k$. G_0 is a subgroup of order q^{k-1} of the group G and for arbitrary $u \neq 0,\ u \in V_1$, the set $K_u = \{\underline{\mathbf{c}}(\underline{\mathbf{x}}') \mid \underline{\mathbf{x}}' \in \mathcal{D}_u\}$ is the coset of G_0. Since the probability of the subgroup G_0' is $q_{k+m}T_0$, by the inductive hypothesis we have

$$\frac{P\left(K_{-v}\right)}{P(G_0')} = \frac{R_{-v}}{T_0} \leq \phi_{k-1}(\underline{\mathbf{p}}) \leq \phi_k(\underline{\mathbf{p}}),$$

see Property 2 of the function $F_k(\underline{\mathbf{p}})$, or

$$T_0\phi_k(\underline{\mathbf{p}}) - R_{-v} > 0.$$

Combining (39) with this inequality, we get the inequality

$$\begin{aligned} p_{k+m}(-\phi_k(\underline{\mathbf{p}})T + R_0 + R - R_{-v}) &\leq p_{k+m}(\phi_k(\underline{\mathbf{p}})T_0 - R_{-v}) \\ &< q_{k+m}(\phi_k(\underline{\mathbf{p}})T_0 - R_{-v}), \end{aligned}$$

which is equivalent to $\phi_k(\underline{\mathbf{p}})P(G) > P(K)$. This completes the proof of (7) for any $m \geq 1$, when the subgroup is of order q^k. Thus, we proved by induction that inequality (7) holds for the subgroups of order q^k, for any $k,\ 0 \leq k < n$.

Proof of Theorem 2

By an appropriate substitution we shall reduce the problem to the case of $\underline{\mathbf{y}} = \underline{\mathbf{0}}$. Define the random variable $\underline{\mathbf{E}}'$ by $\underline{\mathbf{E}}' = \underline{\mathbf{E}} + \underline{\mathbf{x}} + \underline{\mathbf{y}}$. Next, define the vector sequence $\{\underline{\bar{\mathbf{P}}}^{(j)}\}_{j\geq 0}$ by

$$\bar{P}_i^{(j)} = \begin{cases} P_i^{(j)}, & y_i = x_i \\ 1 - P_i^{(j)}, & y_i \neq x_i \end{cases}, \qquad 1 \leq i \leq n,\ j \geq 0. \tag{40}$$

It is evident that $\underline{\mathbf{P}}^{(j)}$ is the error probability vector of $\underline{\mathbf{E}}$ if, and only if, $\bar{\underline{\mathbf{P}}}^{(j)}$ is the error probability vector of $\underline{\mathbf{E}}'$, $j \geq 0$. The sequence $\{\bar{\underline{\mathbf{P}}}^{(j)}\}_{j\geq 0}$ satisfies the recurrent relation

$$\bar{\underline{\mathbf{P}}}^{(j+1)} = \mathcal{F}_{\underline{\mathbf{0}}}(\bar{\underline{\mathbf{P}}}^{(j)}), \qquad j \geq 0,$$

because if for some fixed $j \geq 0$ we denote $\mathcal{F}_{\underline{\mathbf{0}}}(\bar{\underline{\mathbf{P}}}^{(j)})$ by $\underline{\mathbf{P}}$, then

$$\begin{aligned} P_i &= P_{\bar{\underline{\mathbf{P}}}^{(j)}}\Big(\{E_i = 1\} \mid \{\underline{\mathbf{H}}^{(i)}\underline{\mathbf{E}} = \underline{\mathbf{0}}\}\Big) \\ &= P_{\underline{\mathbf{P}}^{(j)}}\Big(\{E_i' = 1\} \mid \{\underline{\mathbf{H}}^{(i)}\underline{\mathbf{E}}' = \underline{\mathbf{0}}\}\Big) \\ &= P_{\underline{\mathbf{P}}^{(j)}}\Big(\{E_i = 1 + x_i + y_i\} \mid \{\underline{\mathbf{H}}^{(i)}\underline{\mathbf{E}} = \underline{\mathbf{H}}^{(i)}\underline{\mathbf{y}}\}\Big) \\ &= \begin{cases} P_i^{(j+1)}, & x_i = y_i \\ 1 - P_i^{(j+1)}, & x_i \neq y_i \end{cases} = \bar{P}_i^{(j+1)}, \qquad 1 \leq i \leq n. \end{aligned}$$

Note that the coordinates of $\bar{\underline{\mathbf{P}}}^{(d)}$ satisfy the inequalities

$$\bar{P}_i^{(d)} < 1/2, \qquad 1 \leq i \leq n, \tag{41}$$

see (40) and (13). If we define the sequence $\{\underline{\mathbf{A}}^{(j)}\}_{j\geq 0}$ by

$$A_i^{(j)} = \frac{1 - \bar{P}_i^{(j)}}{\bar{P}_i^{(j)}} = \frac{1}{\bar{P}_i^{(j)}} - 1, \qquad 1 \leq i \leq n,\ j \geq 0, \tag{42}$$

then from (41) it follows that $A_i^{(d)} > 1$ for all $i,\ 1 \leq i \leq n$.

Denote by C_i^u the set obtained from the set $\{\underline{\mathbf{e}} \in V_n \mid \underline{\mathbf{H}}^{(i)}\underline{\mathbf{e}} = \underline{\mathbf{0}},\ e_i = u\}$ by deleting the i–th coordinate in every element of it, $u \in \{0,1\},\ 1 \leq i \leq n$. The set C_i^0 is a subgroup of order $\exp_2(k_i)$ ($k_i \geq k$) of the group V_{n-1}, and the set C_i^1 is coset of C_i^0, $1 \leq i \leq n$.

Let $\bar{\underline{\mathbf{P}}}^{(j,i)}$ denote the vector obtained from $\bar{\underline{\mathbf{P}}}^{(j)}$ by deleting its i–th coordinate, $j \geq d,\ 1 \leq i \leq n$, and let

$$\delta_i = F_{k_i}(\bar{\underline{\mathbf{P}}}^{(d,i)}) > 1, \qquad 1 \leq i \leq n,$$

see (5). By induction it can be proved that for all $j,\ j \geq d$, we have $A_i^{(j)} > 1,\ 1 \leq i \leq n$, and that for $j > d$ the stronger inequality

$$A_i^{(j)} > A_i^{(j-1)}\delta_i > 1, \qquad 1 \leq i \leq n, \tag{43}$$

holds. For $j = d$ this statement is obviously true. Suppose that it is true for $d, d+1, \ldots, j$. According to Theorem 1, for $j > d$ we have

$$P_{\underline{\bar{\mathbf{P}}}^{(j,i)}}\left(C_i^0\right)/P_{\underline{\bar{\mathbf{P}}}^{(j,i)}}\left(C_i^1\right) \geq F_{k_i}(\underline{\bar{\mathbf{P}}}^{(j,i)}), \qquad 1 \leq i \leq n.$$

By the inductive hypothesis and (42) the coordinates of $\underline{\bar{\mathbf{P}}}^{(j,i)}$ are greater than the corresponding coordinates of $\underline{\bar{\mathbf{P}}}^{(d,i)}$, and so by Property 3 of the function F we have

$$F_{k_i}(\underline{\bar{\mathbf{P}}}^{(j,i)}) \geq F_{k_i}(\underline{\bar{\mathbf{P}}}^{(d,i)}) = \delta_i, \qquad 1 \leq i \leq n.$$

For $j = d$ this inequality is obviously true. From (10) and (42) we have

$$A_i^{(j+1)} = A_i^{(j)} P_{\underline{\bar{\mathbf{P}}}^{(j,i)}}\left(C_i^0\right)/P_{\underline{\bar{\mathbf{P}}}^{(j,i)}}\left(C_i^1\right),$$

and from the last two inequalities, it follows that $A_i^{(j+1)} > A_i^{(j)}\delta_i,\ 1 \leq i \leq n$. Therefore, it is proved by induction that (43) holds for all $j,\ j > d$. As an immediate consequence of (43) we have

$$\lim_{j\to\infty} A_i^{(j)} = +\infty, \qquad 1 \leq i \leq n,$$

and further, because of (42),

$$\lim_{j\to\infty} \bar{P}_i^{(j)} = 0, \qquad 1 \leq i \leq n.$$

The statement of Theorem 2 directly follows from this equation and (40).

Proof of Theorem 3

As in Theorem 2 the derivation is based on a reduction to the case where the received message $\underline{\mathbf{y}}$ equals $\underline{\mathbf{0}}$. Define the matrix sequence $\{\underline{\bar{\mathbf{B}}}^{(s)}\}_{s\geq 0}$ by means of the sequence $\{\underline{\mathbf{B}}^{(s)}\}_{s\geq 0}$ as follows

$$\bar{B}^{(s)}_{r,i} = \begin{cases} B^{(s)}_{r,i}, & y_i = x_i \\ 1 - B^{(s)}_{r,i}, & y_i \neq x_i \end{cases}, \qquad 1 \leq r \leq R,\ 1 \leq i \leq n,\ s \geq 0 \tag{44}$$

Next, define the random variable $\underline{\bar{\mathbf{E}}}$ by $\underline{\bar{\mathbf{E}}} = \underline{\mathbf{E}} + \underline{\mathbf{x}} + \underline{\mathbf{y}}$. Denote by $\underline{\bar{\mathbf{p}}}$ the error probability vector of $\underline{\bar{\mathbf{E}}}$, i.e.

$$\bar{p}_i = \begin{cases} p_i, & x_i = y_i \\ 1 - p_i, & x_i \neq y_i \end{cases}, \qquad 1 \leq i \leq n,$$

where $\underline{\mathbf{p}}$ is the error probability vector of $\underline{\mathbf{E}}$. For an arbitrary $s \geq 0$ and $(r,i) \in \mathcal{H}_1$ (22) let $\underline{\mathbf{g}} = \underline{\mathbf{g}}(\underline{\mathbf{H}}, \underline{\mathbf{B}}^{(s)}, r, i)$ and $\underline{\bar{\mathbf{g}}} = \underline{\mathbf{g}}(\underline{\mathbf{H}}, \underline{\bar{\mathbf{B}}}^{(s)}, r, i)$, see (23). We have

$$\bar{g}_i = \begin{cases} g_i, & x_i = y_i \\ 1 - g_i, & x_i \neq y_i \end{cases}, \qquad 1 \leq i \leq n,$$

and therefore $\underline{\mathbf{g}}$ is the error probability vector of $\underline{\mathbf{E}}$ if, and only if, $\underline{\bar{\mathbf{g}}}$ is the error probability vector of $\underline{\bar{\mathbf{E}}}$.

If we denote $\Phi_{\underline{\mathbf{0}},\underline{\bar{\mathbf{p}}}}(\underline{\bar{\mathbf{B}}}^{(s)})$ by $\underline{\mathbf{B}}$, then for $(r,i) \in \mathcal{H}_1$ we have

$$\begin{aligned} B_{r,i} &= P_{\underline{\bar{\mathbf{g}}}}\left(\{E_i = 1\} \mid \{\underline{\mathbf{H}}^{(i,r)}\underline{\mathbf{E}} = \underline{\mathbf{0}}\}\right) \\ &= P_{\underline{\mathbf{g}}}\left(\{\bar{E}_i = 1\} \mid \{\underline{\mathbf{H}}^{(i,r)}\underline{\bar{\mathbf{E}}} = \underline{\mathbf{0}}\}\right) \\ &= P_{\underline{\mathbf{g}}}\left(\{E_i = 1 + x_i + y_i\} \mid \{\underline{\mathbf{H}}^{(i,r)}\underline{\mathbf{E}} = \underline{\mathbf{H}}^{(i,r)}\underline{\mathbf{y}}\}\right) \\ &= \begin{cases} B^{(s+1)}_{r,i}, & x_i = y_i \\ 1 - B^{(s+1)}_{r,i}, & x_i \neq y_i \end{cases} = \bar{B}^{(s+1)}_{r,i}. \end{aligned}$$

We conclude that the sequence $\{\underline{\bar{\mathbf{B}}}^{(s)}\}_{s\geq 0}$ satisfies the recurrent relation

$$\underline{\bar{\mathbf{B}}}^{(s+1)} = \Phi_{\underline{\mathbf{0}},\underline{\bar{\mathbf{p}}}}(\underline{\bar{\mathbf{B}}}^{(s)}), \qquad s \geq 0.$$

From (28) and (44) it follows that the elements $\bar{B}^{(d)}_{r,i}$, $(r,i) \in \mathcal{H}_1$, of $\underline{\bar{\mathbf{B}}}^{(d)}$ are less than t_0.

For $u \in \{0,1\}$ and $(r,i) \in \mathcal{H}_1$ denote by $C^u_{i,r}$ the set of vectors $\{\underline{\mathbf{e}} \in V_n \mid \underline{\mathbf{H}}^{(i,r)}\underline{\mathbf{e}} = \underline{\mathbf{0}},\ e_i = u\}$. The set $C^0_{i,r}$ is a subgroup of V_n, and the set $C^1_{i,r}$ is a coset of $C^0_{i,r}$. Let us fix the pair $(r,i) \in \mathcal{H}_1$. The matrix $\underline{\mathbf{H}}^{(i,r)}$ has a simple structure, because it contains $j-1$ orthogonal parity checks with k members. We shall find the lower bound on the ratio $P\{\underline{\mathbf{E}} \in C^0_{i,r}\}/P\{\underline{\mathbf{E}} \in C^1_{i,r}\}$ first (directly, not by the use of Theorem 1), where $\underline{\mathbf{E}}$ is an n–dimensional binary random variable with independent coordinates, and $\underline{\mathbf{p}}$ is the error probability vector of $\underline{\mathbf{E}}$ (with coordinates less than 1/2). Here we shall use a more appropriate notation. The coordinates of $\underline{\mathbf{E}}$ contained in the ι–th parity check are denoted by $E_0,\ E_{\iota,1}, \ldots,\ E_{\iota,k-1},\ 1 \leq \iota \leq j-1$, and the corresponding probabilities are denoted by $P\{E_0 = 1\} = u$ and $P\{E_{\iota,\kappa} = 1\} = u_{\iota,\kappa},\ 1 \leq \iota \leq j-1,\ 1 \leq \kappa \leq k-1$. Using the fact that the parity checks are orthogonal, it can easily be deduced [6, Theorem 1] that

$$\frac{P\{\underline{\mathbf{E}} \in C^0_{i,r}\}}{P\{\underline{\mathbf{E}} \in C^1_{i,r}\}} = \frac{1-u}{u}\prod_{\iota=1}^{j-1}\frac{1+\prod_{\kappa=1}^{k-1}(1-2u_{\iota,\kappa})}{1-\prod_{\kappa=1}^{k-1}(1-2u_{\iota,\kappa})}\ .$$

Since this is a decreasing function of $u_{\iota,\kappa}(u_{\iota,\kappa} < 1/2)$ for $1 \leq \iota \leq j-1,\ 1 \leq \kappa < k$, we have

$$\frac{P\{\underline{\mathbf{E}} \in C^0_{i,r}\}}{P\{\underline{\mathbf{E}} \in C^1_{i,r}\}} \geq \frac{1-u}{u}\left(\frac{1+(1-2U)^{k-1}}{1-(1-2U)^{k-1}}\right)^{j-1}, \tag{45}$$

if

$$U \geq \max\{u_{\iota,\kappa} \mid\ 1 \leq \iota \leq j-1,\ 1 \leq \kappa < k\}\,.$$

Here the equality holds if, and only if, $u_{\iota,\kappa} = U,\ 1 \le \iota \le j-1,\ 1 \le \kappa < k$.

Let

$$b^{(s)} = \max\left\{\bar{B}_{r,i}^{(s)} \mid (r,i) \in \mathcal{H}_1\right\}, \qquad s \ge 0. \tag{46}$$

Combining (24) and (45), we get the inequality

$$\begin{aligned} \bar{B}_{r,i}^{(s+1)} &\le \left(1 + \frac{1-\bar{p}_i}{\bar{p}_i}\left(\frac{1+\left(1-2b^{(s)}\right)^{k-1}}{1-(1-2b^{(s)})^{k-1}}\right)^{j-1}\right)^{-1} \\ &= f_{j,k,\bar{p}_i}\left(b^{(s)}\right) \le f_{j,k,1-p}\left(b^{(s)}\right), \qquad (r,i) \in \mathcal{H}_1,\ s \ge 0, \end{aligned}$$

and consequently

$$b^{(s+1)} \le f_{j,k,1-p}\left(b^{(s)}\right), \qquad s \ge 0.$$

From (44) and (28) it follows that $b^{(d)} < t_0(j,k,1-p)$. But then, using (27), we get the inequality

$$b^{(d+1)} \le f_{j,k,1-p}(b^{(d)}) < b^{(d)}.$$

It is seen that (28) is satisfied if we replace d by $d+1$. By induction over s we conclude that

$$b^{(s+1)} \le f_{j,k,1-p}\left(b^{(s)}\right) < b^{(s)}, \qquad s \ge d \tag{47}$$

Since the sequence $\left\{b^{(s)}\right\}$ is positive and decreasing, it converges. Denote by b^* its limit when $s \to +\infty$. From the continuity of $f_{j,k,1-p}(t)$ and (47), it follows that $b^* = f(b^*)$, and so $b^* = 0$. According to (46) we have

$$\lim_{s\to\infty} \bar{B}_{r,i}^{(s)} = 0, \qquad (r,i) \in \mathcal{H}_1.$$

Finally, from (44) we get the equality (29).

References

[1] G. Battail, U.S. Patent 3 805 236, issued Apr. 16, 1974; French patent application 72 00497, Jan. 7, 1972.

[2] G. Battail, M. C. Decouvelaere, P. Godlewski, “Replication Decoding”, *IEEE Trans. Inform. Theory*, vol. IT–25, pp. 332–345, May 1979.

[3] R. B. Blizard, C. C. Korgel, “An Iterative Probabilistic Threshold Decoding Technique”, in *IEEE Nat. Telecom. Conf. Rec.*, Dec. 1972, pp. 13D–1 — 13D–5.

[4] G. C. Clark, Jr., J. B. Cain, *Error–Correction Coding for Digital Communications*, New York: Plenum Press, 1982.

[5] R. G. Gallager, *Low–Density Parity–Check Codes*, Cambridge: MIT Press, 1960.

[6] R. G. Gallager, “Low–Density Parity–Check Codes”, *IEEE Trans. on Inform. Theory*, vol. IT–8, pp. 21–28, Jan. 1962.

[7] C. R. P. Hartmann, L. D. Rudolph, “An Optimum Symbol–by–Symbol Decoding Rule for Linear Codes”, *IEEE Trans. Inform. Theory*, vol. IT–22, pp. 514–517, Sep. 1976.

[8] W. Meier, O. Staffelbach, “Fast Correlation Attacks on Stream Ciphers”, in *Advances in Cryptology, Eurocrypt’88*, C. G. Günter, Ed., Berlin: Springer, 1988, pp. 300–314.

[9] W. W. Peterson, E. J. Weldon, *Error–Correction Codes*, Cambridge: MIT Press, 1982.

[10] G. R. Redinbo, “Inequalities Between the Probability of a Subspace and the Probabilities of its Cosets”, *IEEE Trans. Inform. Theory*, vol. IT–19, pp. 533–536, July 1973.

[11] D. D. Sullivan, “A Fundamental Inequality Between the Probabilities of Binary Subgroups and Cosets”, *IEEE Trans. Inform. Theory*, vol. IT–13, pp. 91–94, Jan. 1967.

Figure 1: Dependence of the estimated successful decoding probability on the BSC error probability for the $(512, 100)$ code of Example 2, when Algorithm P1 includes $1, 2, 3, 5, 7$ and 10 iterations

Figure 2: Dependence of the estimated successful decoding probability on the Gaussian channel SNR for the $(512, 100)$ code of Example 3, when Algorithm P1 includes $1, 2, 3, 5, 7$ and 10 iterations