AN UPPER BOUND ON THE EXPECTED NUMBER OF COMPUTATIONS FOR MAXIMUM LIKELIHOOD DECODING OF LOW–DENSITY CODES

Vladimir B. Balakirsky

Eindhoven University of Technology, Electrical Engineering Department
P.O.Box 513, 5600 MB Eindhoven,The Netherlands
v.b.balakirsky@ele.tue.nl, http://www.mathematik.uni-bielefeld.de/~vbal/

We describe a realization of the maximum likelihood decoding algorithm when the messages are encoded by a low–density code and transmitted over a binary symmetric channel. The algorithm is based on the introduction of a tree structure in space consisting of all possible noise vectors and principles of sequential decoding with the use of a special metric function. We prove an upper bound on the exponent of the expected number of computations in the ensemble of low–density codes and show that it is much less than the exponent for the exhaustive search.

INTRODUCTION AND STATEMENT OF THE PROBLEM

We consider an information transmission system wherein data are encoded by a binary, linear, block, low–density code $\mathcal{C}$ and transmitted over a binary symmetric channel. The code is defined by the parity–check matrix $\mathbf{H}$ having the dimension $mN \times kN$, where m, k, and N are given parameters, which is constructed in such a way that, for all $i = 1, \ldots, m$, the submatrix consisting of rows $(i-1)N+1, \ldots, iN$ is obtained by some permutation of the columns of the matrix $\mathbf{I}_{k,N} = [\,\mathbf{I}_N \ldots \mathbf{I}_N\,]$ of dimension $N \times kN$, where $\mathbf{I}_N$ is the identity $N \times N$ matrix [1]. We will denote the set consisting of $(kN)!$ permutations of indices $1, \ldots, kN$ by Π_{kN} and the matrix obtained using the permutation $\pi \in \Pi_{kN}$ of columns of the matrix $\mathbf{I}_{k,N}$ by $\pi\mathbf{I}_{k,N}$. The code rate satisfies the inequality $R \geq 1 - m/k + (m-1)/(kN)$ that follows from the observation that the vector of length kN consisting of all 1's belongs to the linear space of each submatrix of the parity check matrix consisting of the rows $(i-1)N+1, \ldots, iN$, where $i = 1, \ldots, m$.

Suppose that $\mathbf{y} \in \{0,1\}^{kN}$ is the vector received at the output of a binary symmetric channel. Then the vector $\mathbf{s}_0 = \mathbf{y}\mathbf{H}^{\mathrm{T}} \in \{0,1\}^{mN}$ can be interpreted as the received syndrome, and the decoder has to construct a vector $\hat{\mathbf{e}} \in \{0,1\}^{kN}$ such that $\hat{\mathbf{e}}\mathbf{H}^{\mathrm{T}} = \mathbf{s}_0$. The vectors having this property form the set $\mathbf{y} \oplus \mathcal{C}$. Any element of this set is a valid estimate of the noise vector at the output of the decoder and $\hat{\mathbf{e}}^* = \arg\min_{\hat{\mathbf{e}} \in \mathbf{y} \oplus \mathcal{C}} \mathrm{wt}(\hat{\mathbf{e}})$ is a maximum likelihood estimate, where wt denotes the Hamming weight of a binary vector.

Low–density codes were introduced by Gallager [1]. Investigation of iterative decoding procedures for these codes was continued by Zyablov and Pinsker [2] and other authors. We will present a revised version of [3] and describe a sequential decoding algorithm that allows us to construct a maximal likelihood estimate of the noise vector with lower complexity than that achievable by verification of all binary vectors having the Hamming weights $0, \ldots, \mathrm{wt}(\hat{\mathbf{e}}^*)$.

SEQUENTIAL DECODING ALGORITHM

To apply the sequential decoding technique, we factorize the space $\{0,1\}^{kN}$ by introducing the following tree structure. Let the tree contain nodes associated with all binary vectors of length kN and have $kN+1$ levels numbered $0, \ldots, kN$. Since the probability of any vector being a noise vector is completely determined by its Hamming weight, we put $\binom{kN}{j}$ vectors of the Hamming weight j to level j, where $j = 0, \ldots, kN$. We will deal with sequential updates of a current estimate of the noise vector $\hat{\mathbf{e}}$ by changing one of the bits of $\hat{\mathbf{e}}$ at each step. If the step is interpreted as passing through the edge of a tree, then the tree should be constructed in such a way that any two nodes are connected by the edge only if the Hamming distance between the corresponding vectors is equal to 1. However, the two guidelines above do not generate a tree and we need an additional constraint which prohibits the entering of the same node by several edges. One possibility is to use the lexicographic ordering for this purpose (see Figure 1).

Definition 1: *The error tree is a tree containing 2^{kN} nodes associated with binary vectors of length kN. The edge going from the node associated with the vector $\mathbf{e}$ and leading to the node associated with the vector $\tilde{\mathbf{e}}$ exists if and only if $\tilde{\mathbf{e}} \in \mathcal{E}_1(\mathbf{e})$, where the set $\mathcal{E}_1(\mathbf{e})$ consists of all vectors $\mathbf{e}' \in \{0,1\}^{kN}$ having the following properties :* $\mathrm{wt}(\mathbf{e}') = \mathrm{wt}(\mathbf{e}) + 1$; $\mathrm{wt}(\mathbf{e} \oplus \mathbf{e}') = 1$; *the last bit 1 of the vector $\mathbf{e}'$ is farther than the last bit 1 of the vector $\mathbf{e}$.*

Definition 2: *Let the function*

$$\Gamma(\hat{\mathbf{e}}|\mathbf{s}_0) = \frac{\mathrm{wt}(\hat{\mathbf{e}}\mathbf{H}^{\mathrm{T}} \oplus \mathbf{s}_0)}{m} + \mathrm{wt}(\hat{\mathbf{e}}) \tag{1}$$

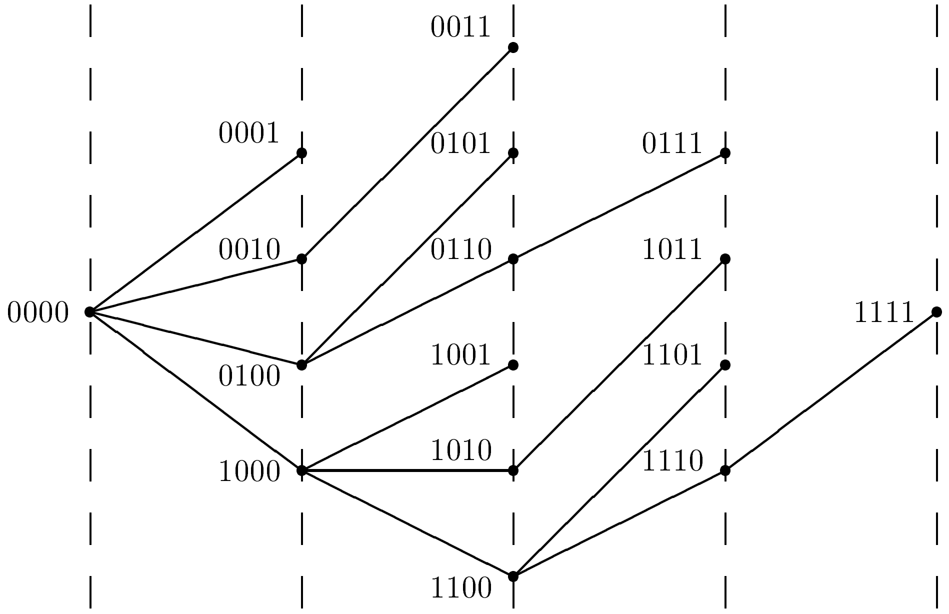

Figure 1: The error tree representing the set $\{0,1\}^{kN}$, where $kN = 4$.

where $\mathbf{s}_0$ *is the received syndrome, be the metric of the node associated with the vector* $\hat{\mathbf{e}}$ *of the error tree.*

One can easily see that the definitions above and the fact that each column of the matrix $\mathbf{H}$ has the Hamming weight m imply the following properties of the metric function : *if* $\mathbf{e}' \in \mathcal{E}_1(\mathbf{e})$, *then* $\Gamma(\mathbf{e}'|\mathbf{s}_0) - \Gamma(\mathbf{e}|\mathbf{s}_0) \in [0,2]$; *if* $\hat{\mathbf{e}}$ *is a valid estimate of the noise vector at the output of the decoder, then* $\Gamma(\hat{\mathbf{e}}|\mathbf{s}_0) = \mathrm{wt}(\hat{\mathbf{e}})$. Consider a curve connecting the points $(\mathrm{wt}(\mathbf{e}), \Gamma(\mathbf{e}|\mathbf{s}_0))$ for the vectors $\mathbf{e}$ associated with the nodes belonging to some path of the error tree. If none of these vectors is a valid estimate of the noise vector at the output of the decoder, then the curve is located strictly above the line with the slope $+1$. All vectors of the set $\mathbf{y} + \mathcal{C}$ specify the corresponding paths in the tree and the corresponding curves. Maximum likelihood decoding is equivalent to the selection of a curve that crosses the line with the slope $+1$ at the point having the minimal abscissa. Suppose that the decoder knows the Hamming weight of the maximum likelihood estimate, $\mathrm{wt}(\hat{\mathbf{e}}^*)$. Then the exponentially lower complexity of the decoding algorithm as compared to the exponent for the exhaustive search is possible only if based on the analysis of the metric of a node associated with a vector $\mathbf{e}$ that does not belong to the path leading to the node associated with the vector $\hat{\mathbf{e}}^*$, the decoder typically can reject all paths containing this node. The properties of the metric function allows us to organize such a verification, namely our decoder rejects these paths if $\Gamma(\mathbf{e}|\mathbf{s}_0) > \mathrm{wt}(\hat{\mathbf{e}}^*)$. However, since the value of $\mathrm{wt}(\hat{\mathbf{e}}^*)$ is unknown to the decoder, he uses the value of a threshold T instead, tries to find a vector $\hat{\mathbf{e}}^*$ with $\Gamma(\hat{\mathbf{e}}^*|\mathbf{s}_0) = \mathrm{wt}(\hat{\mathbf{e}}^*) = T$, and increases T by 1 if there are no vectors having this property. The formal description of the decoding algorithm is given below, where $\mathbf{0}$ denotes the all–zero vector of length kN and $\lceil z \rceil$ stands for the minimal integer which is not less than z.

[I] **Initialization :**
- set $t = 0$, $\mathcal{E}_0 = \{\mathbf{0}\}$, $T = \lceil \Gamma(\mathbf{0}|\mathbf{s}_0) \rceil$.

[S] **Selection of the current node :**
- if $\mathcal{E}_t = \emptyset$, then go to [**B**];
- select an element $\hat{\mathbf{e}}_t$ of the set $\mathcal{E}_t$ having the minimal metric;
- if $\Gamma(\hat{\mathbf{e}}_t|\mathbf{s}_0) = t$, then go to [**T**];
- if $\Gamma(\hat{\mathbf{e}}_t|\mathbf{s}_0) > T$, then go to [**B**].

[F] **The F–step :**
- set $\mathcal{E}_{t+1} = \mathcal{E}_1(\hat{\mathbf{e}}_t)$;
- if $|\mathcal{E}_{t+1}| < T - t$, then go to [**B**];
- increase t by 1 and go to [**S**].

[B] **The B–step :**
- decrease t by 1;
- if $t = 0$, then increase T by 1 and go to [**S**];
- exclude the vector $\hat{\mathbf{e}}_t$ from the set $\mathcal{E}_t$ and go to [**S**].

[T] **Termination :**
- output $\hat{\mathbf{e}} = \hat{\mathbf{e}}_t$ as the estimate of the noise vector.

The statement below directly follows from the description of the algorithm and properties of the metric function.

Proposition : *The* [**I–T**] *sequential decoding algorithm outputs a maximum likelihood estimate of the noise vector.*

AN UPPER BOUND ON THE EXPECTED NUMBER OF COMPUTATIONS

Given $\mathbf{e}_0 \in \{0,1\}^{kN}$ and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$, let

$$C_F(\mathbf{e}_0, \boldsymbol{\pi}) = \Big| \Big\{ \mathbf{e} \in \{0,1\}^{kN} : \Gamma(\mathbf{e}|\mathbf{e}_0(\boldsymbol{\pi}\mathbf{I}_{k,N})^{\mathrm{T}}) \le \Gamma(\mathbf{e}_0|\mathbf{e}_0(\boldsymbol{\pi}\mathbf{I}_{k,N})^{\mathrm{T}}) \Big\} \Big| \quad (2)$$

$$\overline{C}_F(\mathbf{e}_0) = \frac{1}{[\,(kN)!\,]^m} \sum_{\pi_1,\ldots,\pi_m \in \Pi_{kN}} C_F(\mathbf{e}_0, \boldsymbol{\pi}) \quad (3)$$

Table 1: Some values of $\eta_R(\alpha_0)$ and $m_R(\alpha_0)$

α_0	R					R				
	1/4	1/3	1/2	2/3	3/4	1/4	1/3	1/2	2/3	3/4
$\alpha_R/2$	0.50	0.42	0.33	0.27	0.21	3	3	6	9	15
$\alpha_R/4$	0.51	0.40	0.32	0.26	0.21	4	4	6	10	16
$\alpha_R/8$	0.48	0.38	0.30	0.24	0.20	4	5	8	12	19
$\alpha_R/16$	0.47	0.36	0.29	0.23	0.18	5	6	9	14	22
$\alpha_R/32$	0.45	0.35	0.27	0.22	0.17	5	6	10	15	23

where $\mathbf{H} = \boldsymbol{\pi}\mathbf{I}_{k,N}$ denotes the matrix whose submatrix consisting of rows $(i-1)N+1,\ldots,iN$ is constructed as $\pi_i\mathbf{I}_{k,N},\ i=1,\ldots,m$. If $\mathbf{e}_0$ is the noise vector constructed at the output of the decoder, then the algorithm terminates with $T = \mathrm{wt}(\mathbf{e}_0)$. Therefore, all vectors $\mathbf{e}$ with $\Gamma(\mathbf{e}|\mathbf{e}_0\mathbf{H}^{\mathrm{T}}) < \mathrm{wt}(\mathbf{e}_0)$ and some vectors with $\Gamma(\mathbf{e}|\mathbf{e}_0\mathbf{H}^{\mathrm{T}}) = \mathrm{wt}(\mathbf{e}_0)$ will be processed in $[\mathbf{F}]$. If the computation is understood as computing the metric of the vector associated with some node of the error tree, then $TkNC_F(\mathbf{e}_0)$ is an upper bound on the number of computations.

Let $h(p) = -p\log p - (1-p)\log(1-p)$ and $D(p \parallel p') = -p\log p' - (1-p)\log(1-p') - h(p)$ denote the binary entropy function and the divergence between the probability distributions $(p, 1-p)$ and $(p', 1-p')$, respectively, where $p, p' \in (0,1)$ (hereafter, all logarithms are to the base 2). Furthermore, let $D(0 \parallel p') = -\log(1-p')$ and let $|x|^+ = x$ if $x \geq 0$ and $|x|^+ = 0$ if $x < 0$, for all x.

Theorem : *For all* $\mathbf{e}_0 \in \{0,1\}^{kN}$ *such that* $\mathrm{wt}(\mathbf{e}_0) = kN\alpha_0$ *and* $\alpha_0 \in (0, 1/2)$,

$$\frac{1}{kN}\log\overline{C}_F(\mathbf{e}_0) \leq o\left(N^{-1}\log N\right) \tag{4}$$
$$+ \max_{\substack{\alpha\in[0,\alpha_0]\\ \alpha'\in[0,\alpha]}} \left\{ \alpha_0 h\left(\frac{\alpha}{\alpha_0}\right) + (1-\alpha_0)h\left(\frac{\alpha'}{1-\alpha_0}\right) - m\left|\,F^{(k)}(\,k(\alpha-\alpha'), \alpha+\alpha'\,)\,\right|^+ \right\}$$

where

$$F^{(k)}(a,b) = \sup_{\beta:\,\theta_\beta^{(k)} > a} \left[\frac{1}{k}D(\,a \parallel \theta_\beta^{(k)}\,) - D\,(\,b \parallel \beta\,)\right] \tag{5}$$

for all $a, b \in (0,1)$ *and* $\theta_\beta^{(k)} = (1-(1-2\beta)^k)/2$.

Given R and α_0, let $\eta_R(\alpha_0)$ denote the ratio of the function at the right–hand side of (4) and $h(\alpha_0)$, where $k = m/(1-R)$ and the maximum is taken on all integers $m = i(1-R), i = 2, 3, \ldots$ Furthermore, let $m_R(\alpha_0)$ denote the parameter maximizing this function. Then $\eta_R(\alpha_0)$ can be considered as the gain in the exponent of the average number of computations of the decoder as compared to the exhaustive search when α_0 is the fraction of errors in the channel, the codes with the parameters $(m,k) = (m_R(\alpha_0), m_R(\alpha_0)/(1-R))$ are used, and the decoding is correct. Some estimates of $\eta_R(\alpha_0)$ are given in Table 1 where α_R denotes the root of the equation $R = 1-h(\alpha_R)$ corresponding to the Varshamov–Gilbert bound on the minimal distance of a code of rate R. It will follow from the foregoing considerations that, for exponentially many low–density codes, this bound is attained, and the assumption of correct decoding when the fraction of errors is less than $\alpha_R/2$ can be used.

AUXILIARY STATEMENTS

Lemma : *Given* $\mathbf{e} \in \{0,1\}^{kN}$ *and* $w \in \{0,\ldots,mN\}$, *let*

$$\Pr\{\,\mathrm{wt}(\mathbf{eH}^{\mathrm{T}}) \le w\,\} = \frac{|\{\,\pi_1,\ldots,\pi_m \in \Pi_{kN} : \sum_{i=1}^{m} \mathrm{wt}(\mathbf{e}(\pi_i \mathbf{I}_{k,N})^{\mathrm{T}}) \le w\,\}|}{[\,(kN)!\,]^m}$$

If $\mathrm{wt}(\mathbf{e}) = d \in (0, kN/2)$, *then*

$$\frac{1}{kN}\log\Pr\{\,\mathrm{wt}(\mathbf{eH}^{\mathrm{T}}) \le w\,\} \le -mF^{(k)}\left(\frac{w}{mN}, \frac{d}{kN}\right) + m\varepsilon_N\left(\frac{d}{kN}\right) \qquad (6)$$

where $F^{(k)}$ *is the function defined in (5) and* $\varepsilon_N(\lambda) = (2kN)^{-1}\log(2\pi kN\lambda(1-\lambda))$ *for all* $\lambda \in (0,1)$.

Corollary : *For all* $\alpha \in (0,1/2)$,

$$\lim_{N\to\infty}\frac{1}{kN}\log\Pr\{\,\mathrm{wt}(\mathbf{eH}^{\mathrm{T}}) = 0, \ \textit{for some } \mathbf{e} \textit{ with } \mathrm{wt}(\mathbf{e}) = kN\alpha\,\}$$
$$\le \ h(\alpha) - (1-R)\log\frac{1+(1-2\alpha)^k}{2} \qquad (7)$$

i.e., for any $\varepsilon > 0$, *one can find* $k_0(\varepsilon), N_0(\varepsilon) < \infty$ *such that there are exponentially many low-density codes having the parameters* $k \ge k_0(\varepsilon)$, $N \ge N_0(\varepsilon)$, *and the minimal distance* $kN(\alpha_R - \varepsilon)$.

Proof : We will use the lower bound $\binom{kN}{d} \ge 2^{kN(h(\alpha)-\varepsilon_N(\alpha))}$, where $\alpha = d/(kN)$, which follows from Stirling's approximation for the factorial, and the following statement : *let* $d = \mathrm{wt}(\mathbf{e})$ *and* $a_\nu(\mathbf{e}) = |\{\,\pi \in \Pi_{kN} : \mathrm{wt}(\mathbf{e}(\pi\mathbf{I}_{k,N})^{\mathrm{T}}) = \nu\,\}|$

for all $\nu = 0, \ldots, N$. *Then*

$$\frac{a_\nu(\mathbf{e})}{d!(kN-d)!} \le \binom{N}{\nu} \inf_{\beta \in (0,1)} \frac{\left(\theta_\beta^{(k)}\right)^\nu \left(1-\theta_\beta^{(k)}\right)^{N-\nu}}{\beta^d(1-\beta)^{kN-d}} \tag{8}$$

Since $\Pr\{\,\mathrm{wt}(\mathbf{e}\mathbf{H}^{\mathrm{T}}) = 0\,\} = a_0^m(\mathbf{e})/[\,(kN)!\,]^m$, inequality (6) for $w = 0$ directly follows from (8).

If $w > 0$, then we denote the number of permutations $\pi_1, \ldots, \pi_m \in \Pi_{kN}$ such that $\sum_{j=1}^m \mathrm{wt}(\mathbf{e}(\pi_j \mathbf{I}_{k,N})^{\mathrm{T}}) = \mu$ by $A_\mu(\mathbf{e})$, $\mu = 0, \ldots, mN$, introduce a formal variable $z < 0$, and write

$$\left[\sum_{\nu=0}^{N} a_\nu(\mathbf{e}) 2^{\nu z}\right]^m = \sum_{\mu=0}^{mN} A_\mu(\mathbf{e}) 2^{\mu z} \ge 2^{wz}[\,(kN)!\,]^m \Pr\{\,\mathrm{wt}(\mathbf{e}\mathbf{H}^{\mathrm{T}}) \le w\,\}\,.$$

Thus, using (8) and assigning β independent of $\nu \in \{0, \ldots, N\}$, we obtain

$$\begin{aligned}\frac{1}{d!(kN-d)!} \sum_{\nu=0}^{N} a_\nu(\mathbf{e}) 2^{z\nu} \quad &\le \quad \frac{1}{\beta^d(1-\beta)^{kN-d}} \sum_{\nu=0}^{N} \binom{N}{\nu} \left(2^z \theta_\beta^{(k)}\right)^\nu \left(1-\theta_\beta^{(k)}\right)^{N-\nu} \\ &= \quad \frac{\left(1-\theta_\beta^{(k)} + 2^z \theta_\beta^{(k)}\right)^N}{\beta^d(1-\beta)^{kN-d}}\end{aligned}$$

Hence

$$\Pr\{\,\mathrm{wt}(\mathbf{e}\mathbf{H}^{\mathrm{T}}) \le w\,\} \le \inf_{\substack{\beta \in (0,1) \\ z<0}} \left[\frac{d!(kN-d)!}{(kN)!} \cdot \frac{\left(1-\theta_\beta^{(k)} + 2^z\theta_\beta^{(k)}\right)^N}{2^{(w/m)z}\beta^d(1-\beta)^{kN-d}}\right]^m.$$

Given $\beta \in (0,1)$, the expression on the right–hand side is minimized for $z = z_\beta^{(k)}$, where $z_\beta^{(k)} = \log w - \log(mN - w) - \log\theta_\beta^{(k)} + \log(1-\theta_\beta^{(k)})$ and $\theta_\beta^{(k)} > w/(mN)$ implies $z_\beta^{(k)} < 0$. Using these observations and the definition of the divergence, we prove (6) for $w > 0$.

Proof of (8) : Let us denote the column of the matrix $\mathbf{I}_{k,N}$ having 1 in the i–th row by $\mathbf{1}_i$, where $i \in \{1, \ldots, N\}$, and notice that $\mathrm{wt}(\mathbf{e}(\pi\mathbf{I}_{k,N})^{\mathrm{T}}) = \nu$ if and only if there is a vector $(k_1, \ldots, k_N)$ satisfying the constraints

$$k_1, \ldots, k_N \in \{0, \ldots, k\}; \;\; \sum_{i=1}^{N} k_i = d; \;\; |\{i \in \{1, \ldots, N\} : \, k_i \text{ is odd}\,\}| = \nu \tag{9}$$

such that k_1 columns $\mathbf{1}_1, \ldots, k_N$ columns $\mathbf{1}_N$ of the matrix $\pi \mathbf{I}_{k,N}$ are located at the positions where the vector $\mathbf{e}$ contains 1's. Thus

$$\frac{a_\nu(\mathbf{e})}{d!(kN-d)!} = \sum_{(k_1,\ldots,k_N)} \prod_{i=1}^{N} \binom{k}{k_i} \le \binom{N}{\nu} \inf_{s>0} \frac{\left(g_{\mathrm{od}}^{(k)}(s)\right)^\nu \left(g_{\mathrm{ev}}^{(k)}(s)\right)^{N-\nu}}{s^d}$$

where the sum is taken on all vectors $(k_1, \ldots, k_N)$ satisfying (9), s is a formal variable and $g_{\mathrm{od}}^{(k)}(s) = (1+s)^k/2 - (1-s)^k/2$, $g_{\mathrm{ev}}^{(k)}(s) = (1+s)^k/2 + (1-s)^k/2$. We introduce the variable $\beta = s/(1+s)$ and since $g_{\mathrm{od}}^{(k)}(s) = \theta_\beta^{(k)}/(1-\beta)^k$, $g_{\mathrm{ev}}^{(k)}(s) = (1-\theta_\beta^{(k)})/(1-\beta)^k$, express the last inequality as (8).

To prove (7), we use (6), the inequality $\binom{kN}{kN\alpha} \le 2^{kNh(\alpha)}$, and the definition of the divergence $D(\,0 \,\|\, \theta_\alpha^{(k)}\,)$.

PROOF OF THE THEOREM

Substituting (2) to (3) we obtain

$$\overline{C}_F(\mathbf{e}_0) = \sum_{\mathbf{e} \in \{0,1\}^{kN}} \Pr\left\{ \mathrm{wt}((\mathbf{e} \oplus \mathbf{e}_0)\mathbf{H}^{\mathrm{T}}) \le m(\mathrm{wt}(\mathbf{e}_0) - \mathrm{wt}(\mathbf{e})) \right\}. \tag{10}$$

Let $\mathcal{X}_{\alpha,\alpha'}(\mathbf{e}_0)$ be the set consisting of all vectors $\mathbf{e} \in \{0,1\}^{kN}$ such that there are exactly $kN\alpha$ positions j with $(e_{0j}, e_j) = (1,0)$ and exactly $kN\alpha'$ positions j with $(e_{0j}, e_j) = (0,1)$. Then $\mathbf{e} \in \mathcal{X}_{\alpha,\alpha'}(\mathbf{e}_0)$ implies $\mathrm{wt}(\mathbf{e} \oplus \mathbf{e}_0) = kN(\alpha+\alpha')$, $\mathrm{wt}(\mathbf{e}_0) - \mathrm{wt}(\mathbf{e}) = kN(\alpha - \alpha')$, and using (6), we obtain that the logarithm of probability at the right–hand side of (10) divided by kN is upper–bounded by $-mF^{(k)}(k(\alpha - \alpha'), \alpha+\alpha') + m\varepsilon_N\,(\alpha+\alpha')$. Furthermore $|\mathcal{X}_{\alpha,\alpha'}(\mathbf{e}_0)| \le 2^{\alpha_0 h(\alpha/\alpha_0) + (1-\alpha_0)h(\alpha'/(1-\alpha_0))}$ and there are at most $(kN\alpha_0+1)^2/2$ pairs (α, α') such that $\mathcal{X}_{\alpha,\alpha'}(\mathbf{e}_0) \ne \emptyset$. Using these observations and the inequality $\Pr\{\cdot\} \le 1$, we derive (4) from (10).

REFERENCES

[1] R. G. Gallager, *Low–Density Parity–Check Codes.* Cambridge : MIT Press, 1963.

[2] V. V. Zyablov and M. S. Pinsker, “Estimating the complexity of error correction by low–density Gallager codes,” *Probl. Inform. Transmission,* vol. 11, pp. 18–28 (in Russian : vol. 11, no. 1, pp. 23–36), 1975.

[3] V. B. Balakirsky, “Sequential decoding producing the maximum likelihood estimate for the low–density Gallager codes,” *Probl. Inform. Transmission,* vol. 27, pp. 40–47 (in Russian : vol. 27, no. 1, pp. 50–60), 1991.