Information Theory and Signal Processing for (nonlinear) Communication Channels

Konstantinos Xenoulis *

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
kxen@di.uoa.gr

Abstract. In this thesis, error decoding probability bounds and achievable rates for linear and nonlinear communications channels are presented. Gallager's upper bound as well as its variations through the Duman–Salehi bound are improved. The proposed technique relies on the new inverse exponential sum inequality and designates a new desirable characteristic for linear codes, that is directly connected with the concept of list decoding. The thesis also presents lower bounds on the capacity of nonlinear channels represented with Volterra series, combining the random coding technique with the theory of martingales. The proposed research follows the main ideas that dominate Shannon's basic work and properly utilizes exponential martingale inequalities in order to bound the probabilities of erroneous decoding regions. The specific analysis is also applied to cases where the noise statistical characteristics (mean value, deviation) remain unknown. The present work improves and extends the bound of Shulman–Feder for the family of binary, linear codes that are permutation invariant under list decoding. A new upper bound on list error decoding probability is presented that combines random coding techniques for non–random codes and decreases double exponentially with respect to the code's block length.

1 Introduction

Error probability evaluation and capacity are significant performance measures of coded information transmission over various communication channels. The high complexities involved in the calculation of error probability necessitates the introduction of efficient bounding techniques. Classical treatments [1] as well as modern approaches [2] provide tight bounds mostly for random and specific families of codes (turbo codes [3], LDPC codes [4]), since the latter are treated more easily than specific codes. Thus the existence of at least one optimum code within these families is assured, but the respective characteristics of the optimum code remain unknown. The development of new bounding techniques is crucial to the accommodation of optimum specific codes, which can achieve arbitrarily low error decoding probability with rates close to the channel's capacity.

* Dissertation Advisor: Nicholas Kalouptsidis, Professor

This summary is organized as follows; First, Section 2 introduces the basic error bounding techniques and the presents an improvement for discrete channels. Section 3 provides achievable rates for nonlinear channels under maximum likelihood and weakly typical set decoding. A new double exponential upper bound on list error decoding probability is presented in Section 4. It applies to specific codes while combines random coding techniques. Finally, concluding results are given in Section 5.

2 Error probability bounds

Let $\mathcal{C}$ be a block code of length N and dimension k, over a field F with q elements. Let also $\boldsymbol{x}_i \in F^N, i = 0, \ldots, q^k - 1$ and $S_d, d = d_{\min}, \ldots, N$ denote respectively the codewords and the distance distribution of the code $\mathcal{C}$, with $d_{\min}$ its minimum distance. For a vector $\boldsymbol{x} \in F^N$, $wt(\boldsymbol{x})$ denoted its corresponding Hamming weight. For an arbitrary set of messages $\mathcal{M}$ with cardinality M, a message $m,\ 0 \leq m \leq M-1$, is mapped to a codeword $\boldsymbol{x}_m$ of the above code $\mathcal{C}$ and is transmitted over a discrete communication channel with transition probability $P_N(\boldsymbol{y}|\boldsymbol{x}_m)$. $\boldsymbol{y}$ is the received vector at the output of the channel, also of length N. The set of received vectors is denoted by $\boldsymbol{Y}$. Each received vector $\boldsymbol{y}$ is decoded back onto the set of messages $\mathcal{M}$, according to the maximum likelihood (ML) rule. For the aforementioned transmission procedure, Gallager's upper bound [1] on the code's error decoding probability yields

$$P_{e|m} \leq \sum_{\boldsymbol{y} \in \boldsymbol{Y}} P_N(\boldsymbol{y}|\boldsymbol{x}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho) \tag{1}$$

where

$$\mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho) = \left(\sum_{m' \neq m} \left(\frac{P_N(\boldsymbol{y}|\boldsymbol{x}_{m'})}{P_N(\boldsymbol{y}|\boldsymbol{x}_m)} \right)^{\lambda} \right)^{\rho} \tag{2}$$

and $\lambda, \rho \geq 0$. A modified version is provided by the DS2 technique [2, sec. 4.2.2]. Let $G_m(\boldsymbol{y})$ be an arbitrary nonnegative function over $\boldsymbol{Y}$, that may also depend on the transmitted message m. Then, for $\lambda \geq 0$ and $0 \leq \rho \leq 1$,

$$P_{e|m} \leq \left(\sum_{\boldsymbol{y} \in \boldsymbol{Y}} P_N(\boldsymbol{y}|\boldsymbol{x}_m) G_m(\boldsymbol{y}) \right)^{1-\rho} \cdot \left(\sum_{m' \neq m} \sum_{\boldsymbol{y} \in \boldsymbol{Y}} P_N(\boldsymbol{y}|\boldsymbol{x}_m) G_m(\boldsymbol{y})^{1-\frac{1}{\rho}} \left(\frac{P_N(\boldsymbol{y}|\boldsymbol{x}_{m'})}{P_N(\boldsymbol{y}|\boldsymbol{x}_m)} \right)^{\lambda} \right)^{\rho}. \tag{3}$$

The introduction of a tighter upper bound on the ML error decoding probability is made possible by the following inverse exponential sum inequality.

Theorem 1 (Inverse exponential sum inequality [5]). *For positive numbers, $\alpha_1, \alpha_2, \ldots, \alpha_N$ and $\beta_1, \beta_2, \ldots, \beta_N$*

$$\sum_{i=1}^{N} \beta_i \leq \left(\sum_{i=1}^{N} \alpha_i \right) \ln \frac{\sum_{i=1}^{N} \alpha_i e^{\frac{\beta_i}{\alpha_i}}}{\sum_{i=1}^{N} \alpha_i} \tag{4}$$

with equality if and only if $\frac{\beta_i}{\alpha_i}$ =const.

The inverse exponential sum inequality of theorem 1 is used below in the error decoding probability analysis.

Theorem 2 ([5]). *Consider the transmission of an arbitrary set of messages $\mathcal{M}$ over a discrete channel, through the utilization of an (N, R) code $\mathcal{C}$. Let $\boldsymbol{Y}_m^b$ denote the set of erroneous received vectors given that the message m is transmitted*

$$\boldsymbol{Y}_m^b = \left\{ \boldsymbol{y} \in \boldsymbol{Y} : \exists\ m' \in \mathcal{M}, m' \neq m, P_N(\boldsymbol{y}|\boldsymbol{x}_{m'}) \geq P_N(\boldsymbol{y}|\boldsymbol{x}_m) \right\}$$

and

$$\mathcal{L}_m(C, \lambda, \rho) = \min_{\boldsymbol{y} \in \boldsymbol{Y}_m^b} \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho).$$

Then the ML word error decoding probability for the specific code, given that the message m is transmitted, is upper bounded for all $\lambda, \rho \geq 0$ by

$$\begin{aligned} P_{e|m} &\leq \mathcal{L}_m(C, \lambda, \rho)^{-1} \left(\sum_{\boldsymbol{y} \in \boldsymbol{Y}} P_N(\boldsymbol{y}|\boldsymbol{x}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho) \right) \\ &\leq \sum_{\boldsymbol{y} \in \boldsymbol{Y}} P_N(\boldsymbol{y}|\boldsymbol{x}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho). \end{aligned} \tag{5}$$

Theorem 2 provides a bound on the ML word error decoding probability that is tighter than Gallager bound, as noted from the second inequality in (5). Moreover, the DS2 technique can be applied to the second term of the first inequality in (5) for all $\rho \leq 1$, thus leading to a tighter version of the DS2 bound.

Theorem 3 ([5]). *Under the assumptions of theorem 2, the ML word error decoding probability is upper bounded for all $\lambda \geq 0, 0 \leq \rho \leq 1$ and any nonnegative function $G_m(\boldsymbol{y})$ by*

$$\begin{aligned} P_{e|m} &\leq \mathcal{L}_m(C, \lambda, \rho)^{-1} \left(\sum_{\boldsymbol{y} \in \boldsymbol{Y}} P_N(\boldsymbol{y}|\boldsymbol{x}_m) G_m(\boldsymbol{y}) \right)^{1-\rho} \cdot \\ &\left(\sum_{m' \neq m} \sum_{\boldsymbol{y} \in \boldsymbol{Y}} P_N(\boldsymbol{y}|\boldsymbol{x}_m) G_m(\boldsymbol{y})^{1-\frac{1}{\rho}} \left(\frac{P_N(\boldsymbol{y}|\boldsymbol{x}_{m'})}{P_N(\boldsymbol{y}|\boldsymbol{x}_m)} \right)^{\lambda} \right)^{\rho}. \end{aligned} \tag{6}$$

2.1 Special Cases of Theorems 2, 3

Discrete Channels and Coset Based Analysis Since the channel is memoryless and output symmetric, it holds

$$\left(\frac{P_N(\boldsymbol{y}|\boldsymbol{x}^\dagger)}{P_N(\boldsymbol{y}|\boldsymbol{x}_0)}\right)^{\lambda} = \left((q-1)\frac{1-p}{p}\right)^{\lambda\left(wt(\boldsymbol{y})-wt(\boldsymbol{y}-\boldsymbol{x}^\dagger)\right)} \tag{7}$$

where $\boldsymbol{x}_0$ the all zero codeword. Given a specific $\boldsymbol{y}^* \in \mathbf{Y}_0^b$, in analogy to [6, eq.(33)], we define

$$\deg(\boldsymbol{y}^*|\boldsymbol{x}_0) = \left|\left\{\boldsymbol{x}_m \in C, m \neq 0 : wt(\boldsymbol{y}^* - \boldsymbol{x}_m) \leq wt(\boldsymbol{y}^*)\right\}\right| \tag{8}$$

where $\deg(\boldsymbol{y}^*|\boldsymbol{x}_0)$ is the number of codewords whose Hamming distance from $\boldsymbol{y}^*$ is lower than or equal to $wt(\boldsymbol{y}^*)$. The corresponding ratio (7) for each of the above codewords is greater or equal to 1 so that

$$\left(\sum_{m'\neq 0}\left(\frac{P_N(\boldsymbol{y}^*|\boldsymbol{x}_{m'})}{P_N(\boldsymbol{y}^*|\boldsymbol{x}_0)}\right)^{\lambda}\right)^{\rho} \geq \deg(\boldsymbol{y}^*|\boldsymbol{x}_0)^{\rho}. \tag{9}$$

In analogy again to [6, eq.(43)], for every $\boldsymbol{y}^* \in \mathbf{Y}_0^b$,

$$\deg(\boldsymbol{y}^*|\boldsymbol{x}_0) = \sum_{w=1}^{j_t} B_w^t - 1, \quad \left\lceil\frac{d_{\min}}{2}\right\rceil \leq j_t < N \tag{10}$$

where t denotes the coset of $\boldsymbol{y}^*$, $j_t = wt(\boldsymbol{y}^*)$ and B_w^t is the number of words of weight w in the coset t. In contrary to [6, eq.(43)], the term $B_{j_t}^t$ contributes to the sum in the right hand side of (10), since the inequality in (8) is not strict. Moreover, the absence of codeword $\boldsymbol{x}_0$ in the previous definition justifies reducing by 1 the aforementioned sum. Consequently, through (9) and (10),

$$\min_{\boldsymbol{y}^*\in\mathbf{Y}_0^b}\left(\sum_{m'\neq 0}\left(\frac{P_N(\boldsymbol{y}^*|\boldsymbol{x}_{m'})}{P_N(\boldsymbol{y}^*|\boldsymbol{x}_0)}\right)^{\lambda}\right)^{\rho} \geq \min_{t\in\mathcal{T}}\left(\min_{\lceil d_{\min}/2\rceil\leq j_t<N}\sum_{w=1}^{j_t} B_w^t - 1\right)^{\rho} \tag{11}$$

where $\mathcal{T}$ is the set of all cosets t of the code $\mathcal{C}$.

Theorem 4 ([5]). *Under the assumptions of theorem 2, the ML word error decoding probability is upper bounded for all $\lambda \geq 0, 0 \leq \rho \leq 1$ and any nonnegative function $g(y)$ by*

$$P_{e|0} \leq \left(\min_{t\in\mathcal{T}}\left(\min_{\lceil d_{\min}/2\rceil\leq j<N}\sum_{w=1}^{j} B_w^t - 1\right)^{\rho}\right)^{-1}\left(\sum_y \Pr(y|0)g(y)\right)^{N(1-\rho)} \cdot$$
$$\left(\sum_{d=d_{\min}}^{N} S_d\left(\sum_y \Pr(y|0)g(y)^{1-\frac{1}{\rho}}\right)^{N-d}\left(\sum_y \Pr(y|0)^{1-\lambda}\Pr(y|1)^{\lambda}g(y)^{1-\frac{1}{\rho}}\right)^{d}\right)^{\rho}. \tag{12}$$

Example 1. Consider the perfect Hamming code of length 7 with its coset weight distribution depicted in Table [7, p.170 ex. (1)]. Since the minimum distance of the code is 3, all cosets with minimum weight at least $\lceil 1.5 \rceil$ are examined. Then for $j_t = 2, \ldots, 7$,

$$\min_{t \in \mathcal{T}} \left(\min_{2 \leq j_t < 7} \sum_{w=1}^{j_t} B_w^t - 1 \right)^{\rho} = (3 + 1 - 1)^{\rho} . \tag{13}$$

The minimum value is achieved for $j_t = 2$, since for $j_t > 2$, the sum over $\mathcal{T}$ in the right hand side of (11) increases. Actually, since the minimum distance of the code is an odd number, there will always exist a term in the left hand side of (11) strictly greater than one.

3 Achievable rates for nonlinear channels

Random coding theorems and achievable rates for nonlinear additive noise channels are presented in this section both under ML and weakly typical decoding. Consider the transmission of an arbitrary set of messages $\mathcal{M}$ with cardinality M over the nonlinear channel

$$\boldsymbol{y} = D\boldsymbol{x} + \boldsymbol{\nu}. \tag{14}$$

where $\boldsymbol{x}$, $\boldsymbol{y}$ the corresponding input–output sequences of the channel and $\boldsymbol{\nu}$ the noise vector. The nonlinear behavior of the channel is represented by the Volterra system D applied to the channel's input sequence $D\boldsymbol{x}$. The components of the latter vector satisfy

$$[Dx]_i = h_0 + \sum_{j=1}^{d} \sum_{i_1=0}^{\mu} \ldots \sum_{i_j=0}^{\mu} h_j(i_1, \ldots, i_j) x_{i-i_1} \ldots x_{i-i_j} \tag{15}$$

where it holds

$$\|D\boldsymbol{x}\|_{\infty} \leq g_D(\|\boldsymbol{x}\|_{\infty}) \leq g_D(r) \tag{16}$$

and

$$g_D(x) = |h_0| + \sum_{j=1}^{d} \|h_j\| x^j, \quad x \geq 0, \quad \|h_j\| = \sum_{i_1=0}^{\mu} \ldots \sum_{i_j=0}^{\mu} |h_j(i_1, \ldots, i_j)|. \tag{17}$$

In the sequel we assume input causality i.e. $x_i = 0$ for all $i \leq 0$, and that the noise vector $\boldsymbol{\nu}$ is i.i.d gaussian with zero mean and variance σ_{ν}^2. An ML error occurs if, given the transmitted message m and the received vector $\boldsymbol{y}$, another message $m' \neq m$ exists such that

$$\|\boldsymbol{y} - D\boldsymbol{x}_{m'}\|_2^2 \leq \|\boldsymbol{y} - D\boldsymbol{x}_m\|_2^2. \tag{18}$$

Under the random coding setup of Gallager [8, Chap. 5], the average ML error decoding probability $\overline{P}_{e,m}$, given the transmitted message m, satisfies

$$\overline{P}_{e,m} \leq ME\left[\exp\left(-\rho\frac{\|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2}{4\sigma_\nu^2}\right)\right]. \tag{19}$$

Suppose that x_j, x'_j are the j–th components of the corresponding random vectors $\boldsymbol{x}, \boldsymbol{x}'$. Let

$$\emptyset = F_0 \subset F_1 \subset \cdots F_N, \quad F_i = \{x_1, \ldots, x_i, x'_1, \ldots, x'_i\} \tag{20}$$

and

$$\begin{aligned} &Y_i = X_i - X_{i-1},\ 1 \leq i \leq N, \quad X_i = E\left[\|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2 \,|F_i\right] \\ &X_0 = E\left[\|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2\right], \quad X_N = \|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2. \end{aligned} \tag{21}$$

We refer to the sequence $\{Y_i\}_{i=1}^N$ as the martingale difference sequence [9] of the random variable X_N with respect to the joint filter $\{F_i\}_{i=0}^N$ in (20). The mean values appearing in (21) are with respect to all codewords the random variables $\boldsymbol{x}, \boldsymbol{x}'$ can be assigned to. Under the previous setup, we note that

$$\sum_{i=1}^N Y_i = X_N - X_0 = \|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2 - E\left[\|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2\right] \tag{22}$$

and thus (19) is equivalently expressed as

$$\overline{P}_{e,m} \leq M\exp\left(-\frac{\rho}{4\sigma_\nu^2}E\left[\|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2\right]\right)E\left[\exp\left(-\frac{\rho}{4\sigma_\nu^2}\sum_{i=1}^N Y_i\right)\right]. \tag{23}$$

Due to the random coding setup and the independency of the ensemble's codewords, it holds

$$E\left[\|D\boldsymbol{x}-D\boldsymbol{x}'\|_2^2\right] = 2\left(\sum_{j=1}^N E\left[([Du]_j)^2\right] - E\left[[Du]_j\right]^2\right) = 2ND_v \tag{24}$$

where $D_v = E\left[([Du])^2\right] - E\left[[Du]\right]^2$. Finally, combining (23) and (24), we obtain

$$\overline{P}_{e,m} \leq \exp\left(NR - \frac{\rho}{2\sigma_\nu^2}N\mathcal{D}_v\right)E\left[\exp\left(-\frac{\rho}{4\sigma_\nu^2}\sum_{i=1}^N Y_i\right)\right]. \tag{25}$$

3.1 Random Coding Theorem

The development of exponential upper bounds for the mean value in the right hand side of (25) requires bounds on the conditional deviations $dev^+(Y_i)$ and conditional variances $var(Y_i|F_{i-1})$, where according to [9, pp. 24]

$$\begin{aligned} &dev^+(Y_i) = \max_{x_i, x'_i} Y_i \\ &var(Y_i|F_{i-1}) = E\left[(X_i - X_{i-1})^2\,|F_{i-1}\right]. \end{aligned} \tag{26}$$

Appropriate bounds are derived in the lemma that follows.

Lemma 1 ([10]). *Under the assumptions that the components of all codewords* $\boldsymbol{x}_m$*,* $0 \le m \le M-1$ *are mutually independent, and* r *is chosen as in (3), the martingale differences* Y_i *(21) satisfy*

$$dev^{+}(-Y_i) \le 4(\mu+1)g_D(r)^2, \quad var(-Y_i|F_{i-1}) \le 16(\mu+1)^2 g_D(r)^4. \tag{27}$$

The bounds provided by Lemma 1 lead to random coding upper bounds on the average ML error decoding probability. Tighter bounds can be obtained analytically for Volterra systems D of short memory.

Theorem 5 ([10]). *Consider the transmission of an arbitrary set of messages* $\mathcal{M}$ *over a nonlinear Volterra additive gaussian noise channel (14). The components of the noise vector are i.i.d. random variables with* 0 *mean value and variance* σ_ν^2*. For each message* m*,* $0 \le m \le M-1$*, an* N*–length codeword* $\boldsymbol{x}_m$ *is selected from the ensemble* $\boldsymbol{C}$ *of* (N,R) *block codes with probability* Q*, independently from all other codewords, and is transmitted over the channel. If ML decoding is performed at the receiver and the assumptions of Lemma 1 about the codewords' components are valid, then the average error decoding probability* $\overline{P}_e$ *is upper bounded as*

$$\overline{P}_e \le e^{-N\left(E_c(Q,D,\sigma_\nu^2)-R\right)} \tag{28}$$

where

$$E_c(Q,D,\sigma_\nu^2) = \begin{cases} \frac{1}{2\sigma_\nu^2}\mathcal{D}_v - \left(\exp\left(\frac{\kappa}{4\sigma_\nu^2}\right) - 1 - \frac{\kappa}{4\sigma_\nu^2}\right), & \mathcal{D}_v > \frac{\kappa}{2}\left(\exp\left(\frac{\kappa}{4\sigma_\nu^2}\right)-1\right) \\ \frac{1}{\kappa}\left(-2\mathcal{D}_v + (\kappa+2\mathcal{D}_v)\ln\left(1+\frac{2\mathcal{D}_v}{\kappa}\right)\right), & \text{otherwise} \end{cases} \tag{29}$$

and $\kappa = 4(\mu+1)g_D(r)^2$.

Corollary 1. *All rates below* $\max_Q E_c(Q,D,\sigma_\nu^2)$ *(29) are achievable for transmission of information over a nonlinear additive gaussian noise channel under ML decoding.*

3.2 Weakly Typical Set Decoding for Nonlinear Systems

In this section, decoding rules for nonlinear channels are interpreted as concentration measures, and martingale theory is utilized. The analysis can be applied to cases where the channel's transition probability law is generally unknown or a suboptimum decoding algorithm is adopted. The nonlinear model (14) is undertaken using the correlation measure $W(\boldsymbol{x},\boldsymbol{y}) = (D\boldsymbol{x})^T\boldsymbol{y} = (D\boldsymbol{x})^T(D\boldsymbol{x}+\boldsymbol{\nu})$. The input output pair $(\boldsymbol{x},\boldsymbol{y})$ is called weakly ϵ–typical, if

$$W(\boldsymbol{x},\boldsymbol{y}) \ge E_{\Pr(\boldsymbol{x},\boldsymbol{y})}\left[W(\boldsymbol{x},\boldsymbol{y})\right] - N\epsilon. \tag{30}$$

Under the weakly typical decoding rule and the random coding setup, an error occurs given that message m is transmitted, either if codeword $\boldsymbol{x}_m$ is selected form the ensemble $\boldsymbol{C}$ such that $(\boldsymbol{x}_m, \boldsymbol{y})$ does not satisfy (30) or if there exists another message $m' \neq m$ for which $\boldsymbol{x}_{m'}$ is selected independently of $\boldsymbol{x}_m$ such that $(\boldsymbol{x}_{m'}, \boldsymbol{y})$ satisfies (30). Thus, the average error decoding probability, given that m is transmitted, equals

$$\overline{P}_{e,m} = \Pr\left((\boldsymbol{x}_m, \boldsymbol{y}) \text{ not } \epsilon\text{–typical} \bigcup_{m' \neq m} (\boldsymbol{x}_{m'}, \boldsymbol{y}) \ \epsilon\text{–typical}\right) \tag{31}$$

and is upper bounded due to the union bound and the Chernoff bound [8, eq. (5.4.11)] for $\lambda_1, \lambda > 0$ as

$$\begin{aligned} \overline{P}_{e,m} \leq & \, E\left[\exp\left(\lambda_1\left(E\left[W(\boldsymbol{x}, \boldsymbol{y})\right] - N\epsilon - W(\boldsymbol{x}_m, \boldsymbol{y})\right)\right)\right] \\ & + \sum_{m' \neq m} E_{Q(\boldsymbol{x}_{m'})P_N(\boldsymbol{y})}\left[\exp\left(\lambda\left(W(\boldsymbol{x}_{m'}, \boldsymbol{y}) - E[W(\boldsymbol{x}, \boldsymbol{y})] + N\epsilon\right)\right)\right]. \end{aligned} \tag{32}$$

The product probability $Q(\boldsymbol{x}_{m'})P_N(\boldsymbol{y})$ in the innermost term in the right hand side of (32) is a direct consequence of the random coding setup. Indeed, $\boldsymbol{x}_{m'}$ is independent of $\boldsymbol{x}_m$ and consequently of $\boldsymbol{y}$. Noting that $\boldsymbol{x}'_m, \boldsymbol{x}_m$ are dummy variables in the above mean values, (32) satisfies

$$\begin{aligned} \overline{P}_{e,m} \leq & \, E\left[\exp\left(\lambda_1\left(E\left[W(\boldsymbol{x}, \boldsymbol{y})\right] - N\epsilon - W(\boldsymbol{x}, \boldsymbol{y})\right)\right)\right] \\ & + M E_{Q(\boldsymbol{x}')P_N(\boldsymbol{y})}\left[\exp\left(\lambda\left(W(\boldsymbol{x}', \boldsymbol{y}) - E[W(\boldsymbol{x}, \boldsymbol{y})] + N\epsilon\right)\right)\right]. \end{aligned} \tag{33}$$

The following lemma is crucial in the development of exponential martingale inequalities and is used in the proof of the random coding theorem under the weakly typical decoding rule.

Lemma 2 ([10]). *Suppose that all noise samples ν_i, $i \in [1, N]$ are normally distributed $\mathcal{N}(0, \sigma_\nu^2)$. Then for any $\lambda > 0$ and $0 < k < 1$*

$$\begin{aligned} E_{x_i, x'_i}\left[\exp\left(\lambda Y'_i\right)|F'_{i-1}\right] \leq & \exp\left(\frac{\lambda^2}{2k} g_D(r)^2 \sigma_\nu^2\right) \\ & \cdot \left(\frac{1}{2}\exp\left(-\frac{\lambda}{1-k}b'\right) + \frac{1}{2}\exp\left(\frac{\lambda}{1-k}b'\right)\right). \end{aligned} \tag{34}$$

Theorem 6 ([10]). *Let the transmission of an arbitrary set of messages $\mathcal{M}$ over an additive noise nonlinear channel, under the same random encoding setup of Theorem 5. Let also the noise samples be i.i.d. and normally distributed $\mathcal{N}(0, \sigma_\nu^2)$, independent from the channel input. Then, for any $\epsilon, \epsilon_1 > 0$ arbitrarily small constants, the average error decoding probability $\overline{P}_e$ is upper bounded as*

$$\overline{P}_e \leq \epsilon_1 + e^{-N\left(E'_c(Q, D, \sigma_\nu^2) - R\right)} \tag{35}$$

where

$$\begin{aligned} E'_c(Q, D, \sigma_\nu^2) = & \max_{0<k<1} \max_{0<\lambda} \lambda(2\mathcal{D}_v - \epsilon) - \frac{\lambda^2}{2k} g_D(r)^2 \sigma_\nu^2 - \\ & \ln\left(\frac{1}{2}\exp\left(-\frac{\lambda}{1-k}b'\right) + \frac{1}{2}\exp\left(\frac{\lambda}{1-k}b'\right)\right) > 0. \end{aligned} \tag{36}$$

Corollary 2. *Considering the transmission of information over a nonlinear Volterra additive gaussian noise channel, all rates below* $\max_Q E'_c(Q, D, \sigma_\nu^2)$ *(36) are achievable for the weakly typical set decoding rule.*

The tightness of the random coding exponent, given by Theorem 6, depends on lower bounds for the error decoding probability of the form provided in [11], for the specific functions $W(\boldsymbol{x}, \boldsymbol{y})$.

4 Random coding techniques for nonrandom codes

In this section, a new double exponential upper bound on the list error decoding probability of specific classes of codes over binary input symmetric output memoryless channels is derived.

When list decoding is performed at the output of the channel with list size L, the conditional error decoding probability of $\mathcal{C}$, given the transmission of message 0, satisfies

$$P_{e|0,\mathcal{C}}^{\mathcal{L}} = \sum_{\boldsymbol{y} \in \boldsymbol{Y}_{0,\mathcal{C}}^{\mathcal{L}}} P_N(\boldsymbol{y}|\boldsymbol{x}_0, \mathcal{C}) \tag{37}$$

where

$$\boldsymbol{Y}_{0,\mathcal{C}}^{\mathcal{L}} = \left\{ \boldsymbol{y} \in \mathcal{J}^N : \ \exists \{l_i\}_{i=1}^{L}, l_i \neq 0 : P_N(\boldsymbol{y}|\boldsymbol{x}_{l_i}, \mathcal{C}) \geq P_N(\boldsymbol{y}|\boldsymbol{x}_0, \mathcal{C}), \ \forall \ i \in [1, L] \right\}. \tag{38}$$

Moreover, if for $\lambda, \rho \geq 0$ we set

$$\Omega_L(\boldsymbol{y}, \mathcal{C}, \lambda, \rho) = \frac{L^\rho P_N(\boldsymbol{y}|\boldsymbol{x}_0, \mathcal{C})^{\lambda\rho}}{\left(\sum_{m \neq 0} P_N(\boldsymbol{y}|\boldsymbol{x}_m, \mathcal{C})^\lambda\right)^\rho} \tag{39}$$

then, due to the definition in (38), the error decoding probability $P_{e|0,\mathcal{C}}^{\mathcal{L}}$ in (37) is upper bounded as

$$P_{e|0,\mathcal{C}}^{\mathcal{L}} \leq \sum_{\boldsymbol{y} \in \boldsymbol{Y}_{0,\mathcal{C}}^{\mathcal{L}}} P_N(\boldsymbol{y}|\boldsymbol{x}_0, \mathcal{C}) e^{1 - e^{\Omega_L(\boldsymbol{y}, \mathcal{C}, \lambda, \rho) - 1}}. \tag{40}$$

The current work is confined to the following L–list permutation invariant codes.

Definition 1 ([12]). *An* (N, R) *linear binary code* $\mathcal{C}$ *with coset weight distribution matrix* $\boldsymbol{\Gamma}$ *is* L*-list permutation invariant if both the following proporties are satisfied:*
$\mathfrak{L}_1$: *there exists a* $w_{opt} \geq \lceil d_{\min}/2 \rceil$ *such that*

$$L = \min_{\kappa \in [1,K], \boldsymbol{\Gamma}_{\kappa, w_{opt}} \neq 0} \boldsymbol{\Gamma}_{\kappa, w_{opt}} - 1 > 0 \quad \textit{and} \quad \max_{\kappa \in [1,K], w < w_{opt}} \boldsymbol{\Gamma}_{\kappa, w} < L + 1.$$

$\mathfrak{L}_2$: *For all* $\kappa \in [1, K]$*, there exists a* $w_\kappa^L > w_{opt}$ *such that*

$$\boldsymbol{\Gamma}_{\kappa, w_{opt}+1} = \ldots = \boldsymbol{\Gamma}_{\kappa, w_\kappa^L - 1} = 0 \quad \textit{and} \quad \boldsymbol{\Gamma}_{\kappa, w_\kappa^L} \geq L + 1$$

From an L–list permutation invariant code $\mathcal{C}$, we construct an ensemble of codes $\mathcal{E}$ by considering all possible symbol position permutation $N \times N$ matrices $\mathcal{P}$. A position permutation matrix $\mathcal{P}$ has a single 1 in every row and every column and is orthogonal, $\mathcal{P}\mathcal{P}^T = \mathcal{P}^T\mathcal{P} = I$. I is the $N \times N$ identity matrix and $\mathcal{P}^T$ the transpose matrix of $\mathcal{P}$. The lemma provided below is crucial in the derivation of the double exponential bound on the list error decoding probability.

Lemma 3 ([12]). *For an (N, R) binary linear block code $\mathcal{C}$ that is L-list permutation invariant, all codes in the permuted ensemble $\mathcal{E}$ have the same error decoding region $\mathbf{Y}_0^{\mathcal{L}}$.*

Due to the channel symmetry, the average list error decoding probability $P_e^{\mathcal{L}}$ of any code $\mathcal{C}$, over all messages in $\mathcal{M}$, equals $P_{e|0,\mathcal{C}}^{\mathcal{L}}$ [13, Appendix C]. Thus, any bound on $P_{e|0,\mathcal{C}}^{\mathcal{L}}$ is also a bound on $P_e^{\mathcal{L}}$. Moreover, for a L-list permutation invariant code $\mathcal{C}$, in all codes of the ensemble $\mathcal{E}$, message 0 is encoded into the all–zeros vector $\boldsymbol{x}_0$. Thus, $P_N(\boldsymbol{y}|\boldsymbol{x}_0, \mathcal{C}) = P_N(\boldsymbol{y}|\boldsymbol{x}_0)$. Consequently, if we take the average over $\mathcal{E}$ on both sides of (40), then due the error decoding region invariance property stated in lemma 3, we have

$$P_{e|0}^{\mathcal{L}} \leq \sum_{\boldsymbol{y} \in \mathbf{Y}_0^{\mathcal{L}}} P_N(\boldsymbol{y}|\boldsymbol{x}_0) E\left[e^{1-e^{\Omega_L(\boldsymbol{y},\mathcal{C},\lambda,\rho)-1}}\right]. \tag{41}$$

Note that the function $\exp(1 - \exp(x - 1))$ is concave for $0 \leq x \leq 1$ since

$$\frac{de^{1-e^{x-1}}}{dx^2} = e^{-1-e^{-1+x}+x}\left(-e + e^x\right) \leq 0, \text{ for } 0 \leq x \leq 1.$$

Moreover, for any $\boldsymbol{y} \in \mathbf{Y}_0^{\mathcal{L}}$, $\Omega_L(\boldsymbol{y}, \mathcal{C}, \lambda, \rho) \leq 1$. Therefore, application of Jensen's inequality to the right hand side of (41) gives

$$P_e^{\mathcal{L}} \leq \sum_{\boldsymbol{y} \in \mathbf{Y}_0^{\mathcal{L}}} P_N(\boldsymbol{y}|\boldsymbol{x}_0) e^{1-e^{E[\Omega_L(\boldsymbol{y},\mathcal{C},\lambda,\rho)]-1}}. \tag{42}$$

The following technical lemma is useful in the derivation of a closed form upper bound on $P_e^{\mathcal{L}}$.

Lemma 4 ([12]). *The mean value of the double exponent in (42) is lower bounded for all $\rho' \geq 0$ as*

$$E\left[\Omega_L(\boldsymbol{y}, \mathcal{C}, \lambda, \rho)\right] \geq L^{\rho' P}\ P_N(\boldsymbol{y}|\boldsymbol{x}_0)^{\frac{\rho'}{1+\rho'}} \cdot \left[(M-1)^{\rho' P}\left(\sum_{l=0}^{N} b_l \left(\frac{v_l}{b_l}\right)^Q\right)^{\rho' \frac{P}{Q}} \left(\sum_{\boldsymbol{x} \in \mathcal{I}^N} 2^{-N} P_N(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho'}}\right)^{\rho'}\right]^{-1} \tag{43}$$

where

$$b_l = \frac{\binom{N}{l}}{2^N}, \quad v_l = \frac{S_l}{M-1}, \quad 0 \leq l \leq N, \quad \frac{1}{P} + \frac{1}{Q} = 1, \quad P, Q \geq 1. \tag{44}$$

Combining lemma 4 with (42) and passing from $\boldsymbol{Y}_0^{\mathcal{L}}$ to the set of all received vectors $\boldsymbol{Y}$, we get the following theorem.

Theorem 7 ([12]). *Consider an (N, R) binary linear block code $\mathcal{C}$ which is L–list permutation invariant with distance spectrum S_l, $0 \leq l \leq N$ and coset weight distribution matrix $\boldsymbol{\Gamma}$. $\mathcal{C}$ is utilized in the transmission of an arbitrary set of messages $\mathcal{M}$, with cardinality $M = 2^{NR}$, over a binary input, symmetric output discrete memoryless channel. If $p/(q-1)$ is the error transition probability of the channel, then the average list error decoding probability, over all messages in $\mathcal{M}$, $P_e^{\mathcal{L}}$ of $\mathcal{C}$ is upper bounded for all $\rho' \geq 0$ as*

$$P_e^{\mathcal{L}} \leq \sum_{h=0}^{N} \binom{N}{h} (1-p)^{N-h} \left(\frac{p}{q-1}\right)^h \sum_{k=\delta(q-2)h}^{h} \binom{h}{k} (q-2)^{h-k} \exp(1) \cdot$$
$$\cdot \exp\left(-\exp\left(\frac{2^{\rho' N}(M-1)^{-\rho' P} L^{\rho' P}\left((1-p)^{N-h}\left(\frac{p}{q-1}\right)^h\right)^{\frac{\rho'}{1+\rho'}}}{\left(\sum_{l=0}^{N} b_l \left(\frac{v_l}{b_l}\right)^Q\right)^{\rho' \frac{P}{Q}} \mathcal{K}_1(p,q,\rho')^{N-h+k} \mathcal{K}_2(p,q,\rho')^{h-k}} - 1\right)\right) \tag{45}$$

where

$$\mathcal{K}_1(p,q,\rho') = \left((1-p)^{\frac{1}{1+\rho'}} + \left(\frac{p}{q-1}\right)^{\frac{1}{1+\rho'}}\right)^{\rho'}, \quad \mathcal{K}_2(p,q,\rho') = \left(\frac{p}{q-1}\right)^{\frac{\rho'}{1+\rho'}}$$
$$\frac{1}{P} + \frac{1}{Q} = 1, \quad P, Q \geq 1 \quad \textit{and} \quad \delta(q-2) = \begin{cases} 1, \, q=2 \\ 0, \, \textit{otherwise} \end{cases}. \tag{46}$$

We note that the upper bound of theorem 7 fails to reproduce the random coding exponent for an L–list permutation invariant code $\mathcal{C}$, as in [14, Th.1]. Additionally, it does not admit a closed form expression for continuous output channel. Nevertheless, since

$$e^{1-e^{x-1}} \leq \frac{1}{x}, \quad x > 0 \tag{47}$$

(45) is tighter than the generalized version of Shulman-Feder bound in [15, eq.(A17)],[13, Cor.8], Moreover, for L-list permutation invariant codes, application of (47) in (45) provides a new version of the generalized SFB, which nicely complements the one presented in [15, eq.(A17)].

5 Conclusions

This thesis deals with issues regarding reliable and efficient information transmission over linear and nonlinear communication channels. For discrete linear symmetric channels, improved upper bounds are developed under maximum likelihood decoding. Furthermore, double exponential upper bounds on the list decoding error probability of specific codes are presented that combine random coding techniques. Finally, the thesis presents achievable rates for nonlinear channels both under maximum likelihood and weakly typical set decoding, utilizing properly the theory of martingales.

References

1. R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, pp. 3–18, Jan. 1965.
2. I. Sason and S. Shamai, "Performance analysis of linear codes under maximum–likelihood decoding: A tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 3, 2006.
3. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error–correcting coding and decoding," in *Proc. IEEE (ICC'93) Geneva, Switzerland*, May 1993, pp. 1064–1070.
4. R. Gallager, *Low–Density Parity–Check Codes.* Cambridge, MA: MIT Press, 1963.
5. K. Xenoulis and N. Kalouptsidis, "Improvement of Gallager upper bound and its variations for discrete channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4204–4210, Sep. 2009.
6. A. Cohen and N. Merhav, "Lower bounds on the error probability of block codes based on improvements on de caen's inequality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 290–310, 2004.
7. F. MacWilliams and N. Sloane, *The Theory of Error–Correcting Codes.* North Holland, 1983.
8. R. Gallager, *Information Theory and Reliable Communication.* John Wiley and Sons, New York, 1968.
9. C. McDiarmid, "Concentration," *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248, 1998.
10. K. Xenoulis and N. Kalouptsidis, "Random coding theorems for nonlinear channels," *IEEE Trans. Inf. Theory,* accepted under changes, 2010.
11. S. Verdu and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.
12. K. Xenoulis and N. Kalouptsidis, "Double exponential performance bounds for permutation invariant binary linear block codes over symmetric channels," *IEEE Trans. Inf. Theory,* submitted, Dec. 2009.
13. E. Hof, I. Sason, and S. Shamai, "Performance bounds for erasure, list and feedback schemes with linear block codes," Aug. 2010, to appear IEEE Trans. Inf. Theory.
14. N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2101–2104, 1999.
15. S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications (ComSoc & IT joint paper award 2003)," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, 2002.