On the Complexity of Suboptimal Decoding for List and Decision Feedback Schemes

J. Freudenberger and V. Zyablov *

Abstract

We analyze the complexity of list decoding and decision feedback schemes for the binary symmetric channel. Both schemes utilize a sub-optimum list decoding algorithm. If no feedback link is available, then the decoding complexity will asymptotically be bounded by $2^{nR(1-R)}$. The reliability function of the list decoding scheme is equivalent to the sphere-packing bound. In the case of the decision feedback scheme the decoding complexity will be of the order $2^{nR(1-C)}$. With decision feedback the reliability function is lower bounded by $C - R$, where C denotes the channel capacity.

1 Introduction

In this paper we analyze coding schemes for the binary symmetric channel (BSC). The standard coding situation for the BSC is that the encoder selects a codeword from a binary code which corresponds to a particular message. The codeword is then transmitted over the noisy channel. Finally, the receiver tries to infer which message was sent by performing maximum-likelihood decoding. For symmetric channels without feedback near-maximum-likelihood decoding performance can be achieved with significantly reduced complexity. In [1], Evseev showed for the binary symmetric channel that virtually all linear codes of any rate R can be decoded with decoding error probability bounded by twice the error probability of maximum-likelihood decoding. The decoding complexity of his algorithm is of the exponential order $2^{nR(1-R)}$. Note that the known complexity of optimum decoding is of order $2^{\min(R,1-R)n}$.

In contrast to (near) maximum-likelihood decoding, we are interested in the decoding complexity for the following situations: a) The decoder generates a list of potential codewords. b) The decoder has the option of not deciding at all, that is, the decoder may reject its estimate and declare a decoding failure. Option a) is reasonable if the encoder is given redundant data. In this case the decoder may retain some equivocation in his decision.List decoding was first studied by Elias in [2]. Later on it was shown that for a list size which is large but not exponential in n, a list-error exponent equal to the sphere-packing exponent $E_{sp}(R, \epsilon)$ could be obtained [3]. If the receiver has some means to request retransmissions, option b) becomes more suitable than maximum-likelihood decoding. The estimated codeword is only accepted if the decision is sufficiently reliable, otherwise an erasure is declared. Each erasure will result in

*J. Freudenberger is with the Dept. of Telecommunications and Applied Information Theory, University of Ulm, Germany, e-mail: juergen.freudenberger@e-technik.uni-ulm.de; V. Zyablov is with the Institute for Information Transmission Problems, Russian Academy of Science, Moscow, e-mail: zyablov@iitp.ru; This work was supported by DFG (Deutsche Forschungsgemeinschaft) under grant Bo 867/9-1.

a request for more redundancy. Forney proved that an exponent $E_f(R,\epsilon) = E_{sp}(R,\epsilon) + C - R$ is attainable with decision feedback [4].

Most commonly, exponential bounds are derived by averaging the probability of a decoding error over an appropriately chosen ensemble of codes. Evidently, at least one code in the ensemble will have a probability of error that is as small as the ensemble-average. In [5], Gallager introduced a different approach based on the average weight spectrum of the standard ensemble of linear codes. Later one, Blokh and Zyablov showed that codes with *good* weight spectrum exist in this ensemble [6]. Then, they derived lower bounds on the error exponent for particular codes with this weight spectrum. Our bounding technique is based on the Blokh-Zyablov approach, i.e., we consider randomly chosen, but fixed codes. This concept allows us to simultaneously bound the reliability and the decoding complexity. For decoding we utilize a suboptimal list decoding algorithm. This decoding algorithm is a list type generalization of bounded distance decoding. Its suboptimal nature allows to reduce the decoding complexity. In particular, if no feedback link is available the decoding complexity will asymptotically be bounded by $2^{nR(1-R)}$. The obtained error exponent for the list decoding scheme is equivalent to the sphere-packing bound. In the case of the decision feedback scheme the decoding complexity will be of the order $2^{nR(1-C)}$. Thus, depends on the code rate R and the channel capacity C. With decision feedback the reliability function is lower bounded by $C - R$.

We prepare the necessary preliminaries in the next section. In the section 3 we define bounded distance list decoding as a generic decoding mapping. All realizations of this mapping achieve the same performance with respect to decoding errors, but different realizations may have different decoding complexity. As an example we present a realization based on information set decoding. In the following two sections, we derive bounds on decoding error probabilities.

2 Preliminaries

Consider the BSC with crossover probability $\epsilon < 0.5$. Let $\mathbb{F}_2 = \{0,1\}$ be the binary field and let $\mathrm{v}_l \in \mathbb{F}_2$, $r_l \in \mathbb{F}_2$, and $e_l \stackrel{\triangle}{=} \mathrm{v}_l + r_l$ denote the input symbol, the output symbol, and the error of the channel at the lth use, respectively. In the following we consider the transmission of binary n-sequences $\mathbf{v} = (\mathrm{v}_0, \ldots, \mathrm{v}_{n-1})$. Similarly, $\mathbf{r} = (r_0, \ldots, r_{n-1})$, and $\mathbf{e} = (e_0, \ldots, e_{n-1})$ denote the received sequence, and the error sequence, respectively. The error process is memoryless and independent of the channel input. The probability of occurrence of a particular error sequence $\mathbf{e}$ is $P(\mathbf{e}) = \epsilon^{\mathrm{wt}(\mathbf{e})}(1-\epsilon)^{n-\mathrm{wt}(\mathbf{e})}$, where $\mathrm{wt}(\mathbf{e})$ is the number of non-zero positions in $\mathbf{e}$, i.e., the Hamming weight of $\mathbf{e}$. Below we use the mixed entropy function $T_2(x,y) \stackrel{\triangle}{=} -x\log_2 y - (1-x)\log_2(1-y)$. The function $T_2(x,x)$ will be denoted $h_2(x)$ and called the binary entropy function. Let $h_2^{-1}(x)$ denote the unique solution of $h(y) = x$, for $0 \le y \le 1/2$. The relative Gilbert-Varshamov distance is given by $\delta(R) = h_2^{-1}(1-R)$. The capacity of the binary symmetric channel with crossover probability ϵ is $C(\epsilon) = 1 - h_2(\epsilon)$. Note that $\delta(\cdot)$ is the inverse capacity function, i.e., $\delta(C(\epsilon)) = \epsilon$. The sphere-packing exponent is $E_{sp}(R,\epsilon) = T_2(\delta,\epsilon) - 1 + R$. In order to estimate the sum of binomial coefficients we will frequently use the following result [7]. Suppose μn is an integer, where $0 < \mu < 1/2$. Then

$$\sum_{k=0}^{\mu n} \binom{n}{k} \le 2^{nh_2(\mu)} \quad . \tag{1}$$

In particular, we have from (1):

$$\binom{n}{\mu n} \leq 2^{nh_2(\mu)} \quad , \tag{2}$$

where we conclude from the symmetry of the binomial coefficient that (2) holds for $0 \leq \mu \leq 1$.

A binary linear code $\mathcal{C}$ of length n and rate $R = k/n$ is a k-dimensional sub-space of $\mathbb{F}_2^n$. Since a linear code is completely specified by a generator matrix, an ensemble of linear codes may be defined in terms of an ensemble of generator matrices. We consider the set of binary linear codes $\mathbb{E}(n,k)$ generated by all binary $k \times n$ matrices, where we select a particular matrix $\mathbf{G}$ from the ensemble by choosing each digit in the matrix independently and equally likely to be 0 and 1. We will require the following lemma. A similar result is derived in [6]. Let $A(w)$ denote the number of codewords of weight w in the code $\mathcal{C}$.

Lemma 1. *For virtually all codes in the ensemble $\mathbb{E}(n,k)$ we have*

$$\begin{array}{lcll} A(w) & \leq & n^2 2^{-(1-R)n} \binom{n}{w} & \text{for } w \geq \delta(R)n \\ A(w) & = & 0 & \text{for } 0 < w < \delta(R)n \end{array} \quad . \tag{3}$$

3 Bounded Distance List Decoding

We consider a list type generalization of bounded distance decoding. This decoding generates lists of variable size. Let $\mathcal{S}_\rho(\mathbf{r})$ denote the sphere in $\mathbb{F}_2^n$ of radius ρ with center $\mathbf{r}$, where $\mathbf{r}$ is the received sequence.

Definition 1 (Bounded distance list decoding). *For a given linear code $\mathcal{C}$ bounded distance list decoding is a mapping $\psi_L : \mathbb{F}_2^n \rightarrow \mathcal{P}(\mathcal{C})$ defined by*

$$\psi_L(\mathbf{r}) \stackrel{\triangle}{=} \mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C} \quad , \tag{4}$$

where $\mathcal{P}(\mathcal{C})$ denotes the power set of the code $\mathcal{C}$.

In words: The result of the bounded distance list decoding is the set of all codewords which belong to the sphere $\mathcal{S}_\rho(\mathbf{r})$. A decoding failure (erasure $\mathcal{X}$) occurs, if the sphere $\mathcal{S}_\rho(\mathbf{r})$ does not contain any codeword from $\mathcal{C}$, i.e., if $\mid \psi_L(\mathbf{r}) \mid = 0$.

Next, we present a particular algorithm which realizes bounded distance list decoding. This algorithm is a variation of a decoding procedure presented by Dumer in [8] and is based on information set decoding. We use the notation $\mathcal{N}$ for the set $\{0, 1, \ldots, n-1\}$ and call $\mathcal{I} \subseteq \mathcal{N}$ an index set. Let $\mathbf{G} = (\mathbf{g}_0, \ldots, \mathbf{g}_{n-1})$ be a matrix with the n columns $\mathbf{g}_0, \ldots, \mathbf{g}_{n-1}$. By $\mathbf{G}_{[\mathcal{I}]}$ we denote the matrix formed by the columns of $\mathbf{G}$ labeled with all indices from $\mathcal{I}$. Similar, the vector $\mathbf{x}_{[\mathcal{I}]}$ is the vector formed from the corresponding symbols of $\mathbf{x}$.

Definition 2 (Information set). *Let $\mathcal{C}$ be a linear code with generator matrix $\mathbf{G}$ and let $\mathcal{I}$ be an index set with $\mid \mathcal{I} \mid = k$. We call $\mathcal{I}$ an information set if the $k \times k$ sub-matrix $\mathbf{G}_{[\mathcal{I}]}$ has full rank. If $\mathbf{G}_{[\mathcal{I}]}$ has full rank for some set $\mathcal{I}$ with $\mid \mathcal{I} \mid > k$, then $\mathcal{I}(j,s)$ is said to be an information superset.*

If $\mathcal{I}$ is an information set according to definition 2, then any two different codewords disagree on the corresponding k positions. Thus, given the k code symbols corresponding to an information set we can uniquely compute the codeword. This fact can be exploited for decoding. Let $\mathcal{I}(j,s) \stackrel{\triangle}{=} (j, j+1(\mathrm{mod} n), \ldots, j+s-1(\mathrm{mod} n))$ be an index set with $s \geq k$ cyclically consecutive positions starting from position j. For a given generator matrix $\mathbf{G}$ we call the set $L(s) \stackrel{\triangle}{=} \{\mathcal{I}(j,s), j \in \mathcal{N}\}$ the sliding window if all index sets $\mathcal{I}(j,s) \in L(s)$ are information supersets.

Algorithm 1. Sliding window list decoding: Let $L(s) \stackrel{\triangle}{=} \{\mathcal{I}(j,s), j \in \mathcal{N}\}$ be the sliding window of the code $\mathcal{C}$. For decoding we take every subset $\mathcal{I}(j,s)$ and re-encode each sub-block $\mathbf{r}_{[\mathcal{I}(j,s)]} - \mathbf{e}$ for any error pattern $\mathbf{e}$ of length s and weight $\mathrm{wt}(\mathbf{e}) \leq \lfloor \frac{\varrho s}{n} \rfloor$. Every newly re-encoded codeword $\mathbf{v}$ is stored in a list if $\mathrm{dist}(\mathbf{v}, \mathbf{r}) \leq \rho$.

In contrast to Dumer's decoding rule we introduced three modifications: First, we allow different radii ρ. In [8], $\rho = \delta(R)n$ is selected which guarantees near-ML performance for all possible channel conditions. Secondly, we perform list decoding, i.e., we store all codewords which belong to $\mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C}$. Moreover, we explicitly test the reliability of all potential decoding estimates $\mathbf{v}$ by comparing the distance $\mathrm{dist}(\mathbf{v}, \mathbf{r})$ with the preselected radius ρ.

Lemma 2. *Let $L(s)$ be the sliding window for a given linear code from the ensemble $I\!E(n,k)$, then sliding window decoding is a realization of bounded distance list decoding.*

Proof. For any codeword in $\mathcal{C}$ with $\mathrm{dist}(\mathbf{v}, \mathbf{r}) \leq \rho$ we obtain an error pattern $\mathbf{r} - \mathbf{v}$ of weight ρ or less. This error pattern produces for at least on subset $L(j,s) \in L(s)$ a sub-block $\mathbf{e}_{[L(j,s)]}$ of weight $\mathrm{wt}(\mathbf{e}_{[L(j,s)]}) \leq \lfloor \frac{\varrho s}{n} \rfloor$ or less. Consequently, any codeword in $\mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C}$ will be re-encoded during this decoding procedure. □

The following analysis of the decoding complexity is essentially due to Dumer. For a proof of the following lemma see [8].

Lemma 3. *For virtually all binary linear codes in the ensemble $I\!E(n,k)$ the sliding window of length $s = k + \lceil 2 \log_2 n \rceil$ forms n information subsets.*

Theorem 2. *Virtually all codes in $I\!E(n,k)$ can be decoded utilizing bounded distance list decoding with a decoding complexity of the exponential order $2^{n[Rh_2(\varrho)+o(n)]}$.*[1]

Proof. The re-encoding of each information sub-block $\mathbf{r}_{[L(j,s)]} - \mathbf{e}_{[L(j,s)]}$ takes at most $o(n^3)$ operations. For the given $s = k + \lceil 2 \log_2 n \rceil$ the number N of trails satisfies

$$N = n \sum_{e=0}^{\lfloor \frac{\varrho s}{n} \rfloor} \binom{s}{e} \leq n \sum_{e=0}^{\varrho s} \binom{s}{e} \leq n 2^{s h_2(\varrho)} = 2^{n[Rh_2(\varrho)+o(n)]} \quad ,$$

where we have used $\varrho = \rho/n$ and (1). The total number of operations is therefore bound from above by $2^{n[Rh_2(\varrho)+o(n)]}$. However, we may require $L \log L$ additional operations if we wish to order the decoder output list with L elements. Nevertheless, this sorting does not alter the exponential order of the complexity. The decoder output list contains at most $2^{nRh_2(\varrho)}$ sequences. The sorting can therefore be done with $2^{n[Rh_2(\varrho)+o(n)]}$ operations. Taking lemma 2 and lemma 3 into consideration the claim follows. □

[1] $o(\cdot)$ is the usual order notation: If $f(n) \in o(g(n))$, then $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$

4 Error Probability Under Bounded Distance List Decoding

In the following we bound the error probability under bounded distance list decoding. The error event under consideration is the event, that the actually tranmitted codeword is not in the decoder output list $\psi(\mathbf{r})$. First, we bound the erasure probability $P_{\mathcal{X}} \stackrel{\triangle}{=} P(\mid \mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C} \mid = 0)$, that the suboptimal decoding results in an empty list.

Lemma 4. *The erasure probability $P_{\mathcal{X}}$ with bounded distance list decoding satisfies for $\varrho = \rho/n \geq \epsilon$:*

$$P_{\mathcal{X}} \leq 2^{-n[T_2(\varrho,\epsilon)-h_2(\varrho)-o(n)]} \quad . \tag{5}$$

Proof. Without loss of generality we assume that the all-zero codeword has been transmitted. Let $\mathbf{e}$ be an error vector of weight e, hence $\mathbf{r} = \mathbf{e}$. If $e \leq \rho$, then $\mathbf{0} \in \mathcal{S}_\rho(\mathbf{e})$. On the other hand, if $e > \rho$ we may have $\mathbf{v} \in \mathcal{S}_\rho(\mathbf{e})$ for some $\mathbf{v} \neq \mathbf{0}$. Therefore, we bound

$$P_{\mathcal{X}} = P(\mid \mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C} \mid = 0) \leq P(\mathbf{e} \notin \mathcal{S}_\rho(\mathbf{0})) \quad .$$

From the union bound we have:

$$P_{\mathcal{X}} \leq \sum_{e=\rho}^{n} \binom{n}{e} \epsilon^e (1-\epsilon)^{n-e} \quad .$$

Bounding the binomial coefficient using (2) we obtain:

$$P_{\mathcal{X}} \leq \sum_{e=\rho}^{n} 2^{nh_2(e/n)} \epsilon^e (1-\epsilon)^{n-e} \quad .$$

Note that $2^{nh_2(e/n)}\epsilon^e(1-\epsilon)^{n-e}$ has a unique maximum for $e = \varrho n, \varrho \geq \epsilon$. Thus, we have

$$P_{\mathcal{X}} \leq n 2^{nh_2(\varrho)} \epsilon^{n\varrho} (1-\epsilon)^{n(1-\varrho)} \quad ,$$

from which we obtain (5). □

Consider the event of an un-detected error under bounded distance list decoding, i.e., the event that the actually transmitted codeword $\mathbf{v}'$ is not in the decoder output list, but this list is not empty

$$P_u(\mathcal{C}) \stackrel{\triangle}{=} P(\mathbf{v}' \notin \psi(\mathbf{r}), \mid \psi(\mathbf{r}) \mid \geq 1) \quad .$$

Lemma 5. *For almost all codes $\mathcal{C} \in I\!E(n,k)$ the probability $P_u(\mathcal{C})$ of an un-detected error under suboptimal list decoding satisfies:*

$$P_u(\mathcal{C}) \leq 2^{-n[1-R+T(\varrho,\epsilon)-2h(\varrho)-o(n)]} \quad . \tag{6}$$

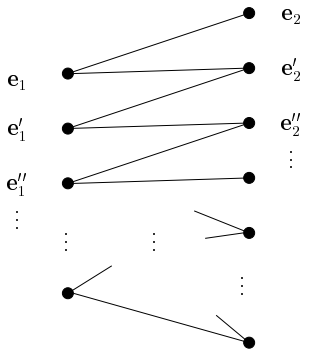

Fig. 1: Graphical representation of lemma 6.

In order to prove lemma 5 we introduce the following quantity: Consider some vector $\mathbf{e}_1$ of weight e_1 and count the number of vectors $\mathbf{e}_2$ which have weight e_2 and distance $\mathrm{dist}(\mathbf{e}_1, \mathbf{e}_2) = l$. Let $U(\mathbf{e}_1, e_2, l)$ denote this number. Note that we can obtain any vector $\mathbf{e}_1'$ of weight e_1 by permuting the binary symbols of $\mathbf{e}_1$. Applying the same permutation to any vector $\mathbf{e}_2$ with $\mathrm{wt}(\mathbf{e}_2) = e_2$ and $\mathrm{dist}(\mathbf{e}_1, \mathbf{e}_2) = l$ we obtain a vector $\mathbf{e}_2'$. This vector satisfies $\mathrm{wt}(\mathbf{e}_2') = e_2$ and $\mathrm{dist}(\mathbf{e}_1', \mathbf{e}_2') = l$. Thus, $U(\mathbf{e}_1, e_2, l)$ is same for all vectors of weight e_1 and is therefore only a function of the weights e_1, e_2, and the distance l. We write $U(e_1, e_2, l)$ for this function. Furthermore, $U(e_1, e_2, l)$ has the following property:

Lemma 6.

$$\binom{n}{e_1} U(e_1, e_2, l) = \binom{n}{e_2} U(e_2, e_1, l) \quad . \tag{7}$$

Proof. Consider the graph in figure 1. Each node on the left side represents a vector of weight e_1, while each node on the right represents a vector of weight e_2. We connect each node from the left with a node from the right by an edge if and only if the distance between the corresponding two vectors is l. Apparently, there are $\binom{n}{e_1} U(e_1, e_2, l)$ edges, because there are $\binom{n}{e_1}$ vectors of weight e_1 on the left each connected to $U(e_1, e_2, l)$ nodes on the right. On the other hand, every node on the right must be connected to $U(e_2, e_1, l)$ nodes on the left. But there are $\binom{n}{e_2}$ vectors of weight e_2 on the right and the proposition follows. □

We will now prove lemma 5.

Proof. We have to bound the probability of an un-detected error with list decoding. That is we are interested in the event $\mathcal{E}$ that the actually transmitted codeword is not in the sphere of radius ρ around the received word $\mathbf{r}$, but there is at least on codeword in this sphere.

We assume without loss of generality that the all-zero codeword has been transmitted. Then, the received sequence is equal to the error vector $\mathbf{e}$. In particular, an un-detected error occurs if the sphere $\mathcal{S}_\rho(\mathbf{e})$ contains at least one codeword $\mathbf{v} \neq \mathbf{0}, \mathbf{v} \in \mathcal{C}$. Yet, we have to consider only error vectors $\mathbf{e}$ with $e = \mathrm{wt}(\mathbf{e}) > \rho$. Consequently, we have

$$P_u(\mathcal{C}) = \sum_{e=\rho+1}^{n} P(\mathcal{E} \mid e) P(e) \quad ,$$

with $P(e) = \binom{n}{e} \epsilon^e (1-\epsilon)^{n-e}$ and

$$P(\mathcal{E} \mid e) = \frac{B(e,\rho)}{\binom{n}{e}} \quad ,$$

where $B(e,\rho)$ denotes the number of error vectors $\mathbf{e}$ of weight e which lead to an un-detected error under list decoding. Thus, we get

$$P_u(\mathcal{C}) = \sum_{e=\rho+1}^{n} B(e,\rho) \epsilon^e (1-\epsilon)^{n-e} \quad . \tag{8}$$

We proceed by bounding the number $B(e,\rho)$. Consider a codeword $\mathbf{v} \in \mathcal{C}$ of weight w. Using the function $U(w,e,l)$ we can estimate the number of possible received vectors $\mathbf{e}$ of weight e such that $\mathbf{v}$ would be in the decoder output list. If $w \leq e$, then there are $\sum_{l=e-w}^{\rho} U(w,e,l)$ such vectors. For $w > e$, we have $\sum_{l=w-e}^{\rho} U(w,e,l)$ vectors. Consequently, summing over all codewords with weights in the range $[e-\rho, e+\rho]$ we can bound $B(e,\rho)$ as follows:

$$B(e,\rho) \leq \sum_{w=e-\rho}^{e} A(w) \sum_{l=e-w}^{\rho} U(w,e,l) + \sum_{w=e+1}^{e+\rho} A(w) \sum_{l=w-e}^{\rho} U(w,e,l) \quad . \tag{9}$$

Substituting the weight spectrum from (3) into (9) we obtain

$$B(e,\rho) \leq n^2 2^{-(1-R)n} \left(\sum_{w=e-\rho}^{e} \binom{n}{w} \sum_{l=e-w}^{\rho} U(w,e,l) + \sum_{w=e+1}^{e+\rho} \binom{n}{w} \sum_{l=w-e}^{\rho} U(w,e,l) \right) \quad .$$

Furthermore, with (7) we have $\binom{n}{w} U(w,e,l) = \binom{n}{e} U(e,w,l)$ and can deduce

$$\begin{aligned} B(e,\rho) &\leq n^2 2^{-(1-R)n} \left(\binom{n}{e} \sum_{w=e-\rho}^{e} \sum_{l=e-w}^{\rho} U(e,w,l) + \binom{n}{e} \sum_{w=e+1}^{e+\rho} \sum_{l=w-e}^{\rho} U(e,w,l) \right) \\ &\leq n^2 2^{-(1-R)n} \left(\binom{n}{e} \sum_{w=e-\rho}^{e+\rho} \sum_{l=0}^{\rho} U(e,w,l) \right) \quad . \end{aligned}$$

Note that $\sum_{l=0}^{\rho} U(e,w,l) \leq \sum_{l=0}^{\rho} \binom{n}{l} \leq 2^{nh(\varrho)}$. Thus, we have

$$B(e,\rho) \leq 2\rho n^2 2^{-(1-R-h(\varrho))n} \binom{n}{e} \quad .$$

Substituting this into (8) we get

$$P_u(\mathcal{C}) \leq 2\rho n^2 2^{-(1-R-h(\varrho))n} \sum_{e=\rho+1}^{n} \binom{n}{e} \epsilon^e (1-\epsilon)^{n-e} \quad .$$

With $2^{-nT(\frac{e}{n},\epsilon)} = \epsilon^e (1-\epsilon)^{n-e}$ and bounding $\binom{n}{e} \leq 2^{nh\left(\frac{e}{n}\right)}$ we obtain

$$P_u(\mathcal{C}) \leq 2\rho n^2 2^{-(1-R-h(\varrho))n} \sum_{e=\rho+1}^{n} 2^{-n(T(\frac{e}{n},\epsilon)-h(\frac{e}{n}))} \quad .$$

The maximum with respect to e is attained for $e = \rho$ with $e \geq \rho \geq \epsilon n$. Thus, we get

$$P_u(\mathcal{C}) \leq 2\rho n^3 2^{-(1-R+T(\varrho,\epsilon)-2h(\varrho))n}$$

which concludes the proof. □

The probability $P_e(\mathcal{C})$ that the actually transmitted codeword is not in the list satisfies

$$P_e(\mathcal{C}) = P_u(\mathcal{C}) + P_{\mathcal{X}} \quad .$$

Choosing $\varrho = \delta$ and minding lemma 5, lemma 4, and theorem 2 we have:

Theorem 3. *For almost all codes $\mathcal{C} \in I\!E(n,k)$ the probability $P_e(\mathcal{C})$ of an error under bounded distance list decoding, that is, the probability that the transmitted codeword is not in the list satisfies:*

$$P_e(\mathcal{C}) \leq 2^{-n[E_{sp}(R,\epsilon)-o(n)]} \quad , \tag{10}$$

where the decoding complexity is of the exponential order $2^{n[R(1-R)+o(n)]}$.

5 Decision Feedback

Consider the decision feedback scheme, where each erasure results in a repeat request for the transmitted codeword. We will now analyze a decision feedback scheme which utilizes bounded distance list decoding. We assume that the receiver chooses the most likely codeword from $\mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C}$ if $\mid \mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C} \mid \geq 1$, else a re-transmission is requested. If the erasure probability is $P_{\mathcal{X}}$, then the probability that a codeword will be repeated i times is $P_{\mathcal{X}}^i$, and the average number of times a codeword is transmitted is

$$1 + P_{\mathcal{X}} + P_{\mathcal{X}}^2 + \ldots = \frac{1}{1-P_{\mathcal{X}}} \quad .$$

Consequently, if the rate of the code is R, the effective rate of information transmission is reduced to

$$R_e = R(1-P_{\mathcal{X}}) \quad . \tag{11}$$

We refer to

$$E_{\mathcal{X}}(R,\epsilon,\varrho) \stackrel{\triangle}{=} \lim_{n\to\infty} -\frac{1}{n}\log_2(P_{\mathcal{X}}) \tag{12}$$

as the erasure exponent. For $E_{\mathcal{X}}(R,\epsilon,\varrho) > 0$ the effective rate can be made as close to R as desired by increasing the codeword length n. Consider the probability of an (un-detected) error with this decision feedback scheme, i.e., $P_{df}(\mathcal{C}) \stackrel{\triangle}{=} P(\text{decoding error} \mid \mid \mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C} \mid \geq 1)$. Above discussion motivates the definition of the feedback exponent $E_{fs}(R)$ as

$$E_{fs}(R) \stackrel{\triangle}{=} \lim_{n\to\infty} -\frac{1}{n}\log_2(P_{df}(\mathcal{C})) \quad , \tag{13}$$

where ϱ is chosen as the limiting value such that $E_{\mathcal{X}}(R,\epsilon,\varrho)$ approaches zero. With (5) we have

$$E_{\mathcal{X}}(R,\epsilon,\varrho) \geq T_2(\varrho,\epsilon) - h_2(\varrho) \quad . \tag{14}$$

That, is $\varrho = \epsilon$ is the limiting value such that $E_{\mathcal{X}}(R,\epsilon,\varrho)$ approaches zero.

Theorem 4. *For almost all codes $\mathcal{C} \in I\!E(n,k)$ the probability $P_{df}(\mathcal{C})$ of an error with suboptimal decoding and decision feedback satisfies:*

$$P_{df}(\mathcal{C}) \leq 2^{-n\left[E_{fs}(R,\epsilon)+o(n)\right]} \tag{15}$$

with

$$E_{fs}(R,\epsilon) \geq \; C(\epsilon) - R \quad \textit{for } R < C \textit{ and } \epsilon < \delta \quad . \tag{16}$$

Moreover, R_e converges to R for $n \to \infty$. The decoding complexity is bounded from above by $2^{n[R(1-C(\epsilon))+o(n)]}$.

Proof. Again, we assume without loss of generality that the all-zero codeword has been transmitted, i.e., $\mathbf{r} = \mathbf{e}$. Let e denote the weight of the error vector $\mathbf{e}$. We can split the event that a decoding error occurs conditioned on $\mid \mathcal{S}_\rho(\mathbf{r}) \cap \mathcal{C} \mid \geq 1$ into two disjoint events: event $\mathcal{E}_1$ that an un-detected error occurs given $e \leq \rho$, and event $\mathcal{E}_2$ that an un-detected error occurs given $e > \rho$. Thus,$P_{df}(\mathcal{C}) \in \mathcal{S}_\rho) = P(\mathcal{E}_1) + P(\mathcal{E}_2)$. First, consider event $\mathcal{E}_1$. Note that $e \leq \rho$ implies $\mathbf{0} \in \mathcal{S}_\rho(\mathbf{r})$. Let $\mathbf{v} \in \mathcal{C}$ be a codeword of weight w. Assume $e = i + l \geq \lceil w/2 \rceil$ with i the number of ones within $\mathrm{supp}(\mathbf{v})$. If the error pattern $\mathbf{e}$ leads to a decoding error, then the distance $\mathrm{dist}(\mathbf{v},\mathbf{e}) = w - i + l \leq e$. Estimating the decoding error probability by the probability of the inequality for the distances, and summing over all code vectors of weight at most 2ρ, we obtain

$$P(\mathcal{E}_1) \leq \sum_{w=d_{min}}^{2\rho} A(w) \sum_{i=\lceil w/2 \rceil}^{\rho} \sum_{l=0}^{\rho-i} \binom{w}{i} \binom{n-w}{l} \epsilon^{i+l} (1-\epsilon)^{n-i-l} \quad . \tag{17}$$

Substituting the weight spectrum from (3) into (17) we obtain

$$P(\mathcal{E}_1) \leq \sum_{w=\delta n}^{2\rho} n^2 2^{-(1-R)n} \binom{n}{w} \sum_{i=\lceil w/2 \rceil}^{\rho} \sum_{l=0}^{\rho-i} \binom{w}{i} \binom{n-w}{l} \epsilon^{i+l} (1-\epsilon)^{n-i-l} \quad .$$

With $w = \omega n$, $\epsilon^{i+l}(1-\epsilon)^{n-i-l} = 2^{-nT_2\left(\frac{i+l}{n},\epsilon\right)}$, and using (2) several times, we have

$$P(\mathcal{E}_1) \leq n^2 \sum_{w=\delta n}^{2\rho} \sum_{i=\lceil w/2 \rceil}^{\rho} \sum_{l=0}^{\rho-i} 2^{-n\left[1-R-h_2(\omega)-\omega h_2\left(\frac{i}{\omega n}\right)-(1-\omega)h_2\left(\frac{l}{n(1-\omega)}\right)+T_2\left(\frac{i+l}{n},\epsilon\right)\right]} \quad .$$

We obtain the maximum over l for $l = (n-w)\epsilon$ if $(n-w)\epsilon$ is within the summation range. However, for $\rho = \varrho n$, with $\varrho \leq \epsilon + (\frac{1}{2} - \epsilon)\delta$, we have $(n-w)\epsilon \geq \rho - i$. Therefore, the maximum is attained for $l = \varrho n - i$. we get

$$P(\mathcal{E}_1) \leq \epsilon n^3 \sum_{w=\delta n}^{2\varrho n} \sum_{i=\lceil w/2 \rceil}^{\varrho n} 2^{-n\left[1-R-h_2(\omega)-\omega h_2\left(\frac{i}{\omega n}\right)-(1-\omega)h_2\left(\frac{\varrho n-i}{n(1-\omega)}\right)+T_2(\varrho,\epsilon)\right]} \quad .$$

Furthermore, we obtain the maximum over i for $i = \varrho w$. But by assumption we have $w/2 \geq \varrho w$ and we choose $i = w/2$:

$$P(\mathcal{E}_1) \leq \epsilon^2 n^4 \sum_{w=\delta n}^{2\varrho n} 2^{-n\left[1-R-h_2(\omega)-\omega-(1-\omega)h_2\left(\frac{\varrho-\omega/2}{1-\omega}\right)+T_2(\varrho,\epsilon)\right]} \quad .$$

Similarly, the maximum of the sum over w is attained for $\omega = 2\varrho(1-\varrho)$. Substituting this into above formula, choosing the limiting values $\varrho = \epsilon$, and simplifying we get

$$P(\mathcal{E}_1) \leq 2^{-n[C-R-\frac{1}{n}\log_2(\epsilon^3 n^5)]} \quad .$$

Now, consider $\mathcal{E}_2$. This is the event of an un-detected error under suboptimal list decoding. With lemma 5 we have

$$P(\mathcal{E}_2) \leq 2^{-n[1-R+T(\varrho,\epsilon)-2h(\varrho)-o(n)]}$$

and $P(\mathcal{E}_2) \leq 2^{-n[C-R-o(n)]}$ in the limit $\varrho \to \epsilon$.

Finally, since for $\varrho > \epsilon$ the rate loss due to re-transmissions remains negligible, R_e converges to R. Moreover, it follows from Theorem 2 that in the limit $\varrho \to \epsilon$ the total number of decoding operations is bounded from above by $2^{n[R(1-C(\epsilon))+o(n)]}$. □

References

[1] G. S. Evseev, "On the complexity of decoding linear codes," *Probl. Pered. Inform.(in Russian)*, vol. 19, pp. 3–8, 1983.

[2] P. Elias, "List decoding for noisy channels," *IEEE Trans. Inform. Theory*, vol. IT-2, pp. 94–104, Apr. 1957.

[3] P. M. Ebert, *Error bounds for parallel communication channels*, M.I.T. Research Laboratory of Electronics, Rept. 448, Cambridge, Massachusetts,, 1966.

[4] Jr. G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 206–220, Mar. 1968.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*, M.I.T. Press, Cambridge, Massachusetts,, 1963.

[6] E. L. Blokh and V. V. Zyablov, *Linear Concatenated Codes*, Nauka, Moscow, 1982, in Russian.

[7] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, North-Holland, 1978.

[8] I. Dumer, "Suboptimum decoding of linear codes: Partition technique," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1971–1986, 1996.