Achievable Error Exponents for Channel with Side Information — Erasure and List Decoding

Erez Sabbag and Neri Merhav

Electronics
Computers
Communications

Achievable Error Exponents for Channel with Side Information – Erasure and List Decoding

Erez Sabbag and Neri Merhav

March 7, 2009

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
{erezs@tx, merhav@ee}.technion.ac.il

Abstract

We consider a decoder with an erasure option and a variable size list decoder for channels with non-casual side information at the transmitter. First, universally achievable error exponents are offered for decoding with an erasure option using a parameterized decoder in the spirit of Csiszár and Körner's decoder. Then, the proposed decoding rule is generalized by extending the range of its parameters to allow variable size list decoding. This extension gives a unified treatment for erasure/list decoding. Exponential bounds on the probability of list error and the average number of incorrect messages on the list are given. Relations to Forney's and Csiszár and Körner's decoders for discrete memoryless channel are discussed. These results are obtained by exploring a random binning code with conditionally constant composition codewords proposed by Moulin and Wang, but with a different decoding rule.

1   Introduction

A decoder with an erasure option is a decoder which has the option of not deciding, i.e., to declare an "erasure". On the other hand, a variable size list decoder is a decoder which produces a list of estimates for the correct message rather than a single estimate, where a list error occurs when the correct message is not on the list. In [1], Forney explored the random coding error exponents of erasure/list decoding for discrete memoryless channels (DMC's). These bounds were obtained by analyzing the optimal decoding rule [1, eq. (11)]

$$\boldsymbol{y} \in \mathcal{R}_m \text{ iff } \Pr(\boldsymbol{y}, \boldsymbol{x}_m) \geq e^{NT} \sum_{m' \neq m} \Pr(\boldsymbol{y}, \boldsymbol{x}_{m'}) \tag{1}$$

where $\Pr(\boldsymbol{y}, \boldsymbol{x}_m)$ is the joint probability of the channel output $\boldsymbol{y}$ and the codeword $\boldsymbol{x}_m$, and T is an arbitrary parameter. The bounds were obtained using Gallager's bounding techniques. Forney showed that the list option and the erasure option are "two sides of the same coin", namely, by changing the value of T one can switch from list decoding (T is negative) to decoding with an erasure option (T is positive).

In [2, Th. 5.11], Csiszár and Körner derived universally achievable error exponents for a decoder with an erasure option for DMC's. These error exponents were obtained by analyzing the following universal decoding rule [2, p. 176] for constant composition (CC) codes:

$$\varphi(\boldsymbol{y}) = \begin{cases} m & , \quad I(\boldsymbol{x}_m; \boldsymbol{y}) > \tilde{R} + \lambda |I(\boldsymbol{x}_{m'}; \boldsymbol{y}) - R|^+ \quad \forall m' \neq m \\ 0 & , \quad \text{else} \end{cases} \tag{2}$$

where R is the code rate (i.e., $m \in \{1, \ldots, 2^{NR}\}$), $\boldsymbol{x}_m$ is a codeword taken from a given type class $T_{\boldsymbol{x}}$, $\boldsymbol{y}$ is the channel output, $I(\boldsymbol{x}; \boldsymbol{y})$ is the empirical mutual information, and $\tilde{R} \geq R$ and $\lambda > 0$ are arbitrary parameters. This decoding rule generalizes the maximum mutual information (MMI) decoder [2, p. 164] to include an erasure option. The bounds were obtained using a fixed composition coding and by applying the packing lemma derived in [2, Lemma 5.1]. However, these bounds were not extended to variable size list decoding. We note that the decoding rule (2) depends on the coding rate R, which might limit its generality. Moreover, it was stated that (2) is an unambiguous decoding rule for $\lambda > 0$, a fact that was used to derive the error exponents. It turns out that this decoding rule is unambiguous only when $\lambda \geq 1$. Unlike Forney's decoder (1), no optimality claims were made for this decoder but, in [3, Sec. 4.4.3] Teletar stated that these bounds are "essentially the same as those in [1]".

Recently, Moulin [4] generalized Csiszár's decoder using a weighting function:

$$\varphi(\boldsymbol{y}) = \begin{cases} m \, , \, I(\boldsymbol{x}_m; \boldsymbol{y}) > R + \max_{m' \neq m} F\big(I(\boldsymbol{x}_{m'}; \boldsymbol{y}) - R\big) \\ 0 \, , \, \text{else} \end{cases} ,$$

where $F(\cdot)$ is a continuous, non-decreasing function. The corresponding error exponents were analyzed and it it was shown that for some rates and channels these error exponents coincide with Forney's error exponents. Note that Moulin's proposed decoder is a function of the code rate R similarly to Csiszár's decoder.

In [5],[6] Teletar and Gallager proposed tighter exponential bounds on decoding with an erasure option and list decoding for DMC using the method of types. These bounds

are not universal in general since the decoding metric depends on the channel statistics. However, it is claimed that under certain conditions these bounds are tighter than Forney's bounds. See [6, Sec.III].

As far as we know, no similar bounds were ever offered for discrete memoryless channels with random states, which are observed by the encoder but not by the decoder [7]. For ordinary decoding (without erasure/list option), Moulin and Wang [8] recently derived an achievable error exponent for channels with state information present non-causally at the transmitter. These results were obtained by analyzing the error probability of a stacked binning scheme and a maximum penalized mutual information (MPMI) decoder.

In this work, we use the random code construction proposed by Moulin and Wang [8] to derive achievable error exponents for decoding with an erasure option and variable size list decoding. In Section 2, we propose a parameterized decoding rule with an erasure option in the spirit of (2). In Section 3, we derive universally achievable error exponents by analyzing the proposed decoding rule. In Section 4, achievable error exponents are offered to decoder with a list option. These exponents are obtained by extending the range of the proposed decoder's parameters to allow decoding with a list option. The generalized decoding rule enables a unified treatment for erasure/list decoding similar to Forney's decoder (1). In Section 5, relations to Forney's and Csiszár and Körner's decoders for DMC are discussed. Moreover, it is shown that the obtained error exponents generalize some known results.

2 Notation and Preliminaries

We begin with some notations and definitions. Throughout this work, capital letters represent scalar random variables (RVs), and specific realizations of them are denoted by the corresponding lowercase letters. Random vectors of dimension N will be denoted by bold-face letters. The notation $\mathbb{1}\{A\}$, where A is an event, will designate the indicator function of A (i.e.,$\mathbb{1}\{A\} = 1$ if A occurs and $\mathbb{1}\{A\} = 0$ otherwise). The notion $a_n \doteq b_n$, for two positive sequences $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$, expresses asymptotic equality in the logarithmic scale, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \ln\left(\frac{a_n}{b_n}\right) = 0.$$

Let the vector $\hat{P}_{\boldsymbol{x}} = \{\hat{P}_{\boldsymbol{x}}(a),\ a \in \mathcal{X}\}$ denote the empirical distribution induced by a vector $\boldsymbol{x} \in \mathcal{X}^n$, where $\hat{P}_{\boldsymbol{x}}(a) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i = a\}$. The type class $T_{\boldsymbol{x}}$ is the set of vectors $\tilde{\boldsymbol{x}} \in \mathcal{X}^n$ such that $\hat{P}_{\tilde{\boldsymbol{x}}} = \hat{P}_{\boldsymbol{x}}$. A type class induced by the empirical distribution $\hat{P}_{\boldsymbol{x}}$ will be denoted by $T(\hat{P}_{\boldsymbol{x}})$. Similarly, the joint empirical distribution induced by $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ is the vector $\hat{P}_{\boldsymbol{xy}} = \left\{\hat{P}_{\boldsymbol{xy}}(a,b),\ a \in \mathcal{X},\ b \in \mathcal{Y}\right\}$ where

$$\hat{P}_{\boldsymbol{xy}}(a,b) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i = a, y_i = b\}, \quad x \in \mathcal{X},\ y \in \mathcal{Y},$$

i.e., $\hat{P}_{\boldsymbol{xy}}(a,b)$ is the relative frequency of the pair (a,b) along the pair sequence $(\boldsymbol{x}, \boldsymbol{y})$. Likewise, the type class $T_{\boldsymbol{xy}}$ is the set of all pairs $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in \mathcal{X}^n \times \mathcal{Y}^n$ such that $\hat{P}_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{y}}} = \hat{P}_{\boldsymbol{xy}}$. The conditional type class $T_{\boldsymbol{y}|\boldsymbol{x}}$, for given vectors $\boldsymbol{x} \in \mathcal{X}^n$, and $\boldsymbol{y} \in \mathcal{Y}^n$ is the set of all vectors $\tilde{\boldsymbol{y}} \in \mathcal{Y}^n$ such that $T_{\boldsymbol{x}\tilde{\boldsymbol{y}}} = T_{\boldsymbol{xy}}$. The Kullback-Leibler divergence between two distributions P and Q on $\mathcal{A}$, where $|\mathcal{A}| < \infty$ is defined as

$$\mathcal{D}(P\|Q) = \sum_{a \in \mathcal{A}} P(a)\ln\frac{P(a)}{Q(a)},$$

with the conventions that $0\ln 0 = 0$, and $p\ln\frac{p}{0} = \infty$ if $p > 0$. We denote the empirical entropy of a vector $\boldsymbol{x} \in \mathcal{X}^n$ by $\hat{H}(\boldsymbol{x})$, where $\hat{H}(\boldsymbol{x}) = -\sum_{a \in \mathcal{X}} \hat{P}_{\boldsymbol{x}}(a)\ln\hat{P}_{\boldsymbol{x}}(a)$. Other information theoretic quantities governed by empirical distributions (e.g., conditional empirical entropy, empirical mutual information) will be denoted similarly. Finally, we define $|t|^+ \triangleq \max\{0, t\}$ and $\exp_2(t) \triangleq 2^t$.

Consider a discrete memoryless state dependent channel with a finite input alphabet $\mathcal{X}$, a finite state alphabet $\mathcal{S}$, a finite output alphabet $\mathcal{Y}$, and a probability transition distribution $W(y|x,s)$. Given an input sequence $\boldsymbol{x}$ and a state sequence $\boldsymbol{s}$ emitted from a discrete memoryless source $P_S(\boldsymbol{s}) = \prod_{i=1}^{N} P_S(s_i)$, the channel output sequence $\boldsymbol{y}$ is generated according to the conditional distribution $W(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}) = \prod_{i=1}^{N} W(y_i|x_i, s_i)$. A message $m \in \{1, \ldots, M\}$ is to be transmitted to the receiver. We assume that the state sequence $\boldsymbol{s}$ is available at the transmitter non-causally, but not at the receiver. We also assume that all messages are a-priori equiprobable. Given $\boldsymbol{s}$ and m, the transmitter produces a sequence $\boldsymbol{x} = f_N(\boldsymbol{s}, m)$ which is used to convey message m to the decoder.

2.1 Codebook construction [8]

In [8, p. 1337], Moulin and Wang used in their derivation a binning code with conditionally constant composition (CCC) codewords. This code will be used in our proofs. For the sake of completeness, we briefly describe the code construction and the encoding process. The decoding part will be described in detail later. The code construction requires the use of an auxiliary random variable $U \in \mathcal{U}$ which takes on values in a finite set of size $|\mathcal{X}||\mathcal{S}| + 1$. See [8, Sec. III.E] for more information.

For a given empirical conditional distribution $\hat{P}^*_{\boldsymbol{xu}|\boldsymbol{s}}$, a sub-code $\mathcal{C}(\hat{P}_{\boldsymbol{s}})$ is constructed for each state sequence type class $T_{\boldsymbol{s}} = T(\hat{P}_{\boldsymbol{s}})$. Given a state type class $T(\hat{P}_{\boldsymbol{s}})$, compute the marginal distribution

$$\hat{P}^*_{\boldsymbol{u}}(u) = \sum_x \sum_s \hat{P}^*_{\boldsymbol{xu}|\boldsymbol{s}}(x, u|s)\hat{P}_{\boldsymbol{s}}(s)$$

where $\hat{P}_{\boldsymbol{s}}$ is the empirical distribution induced $T_{\boldsymbol{s}}$. Note that $\hat{P}^*_{\boldsymbol{u}}(u)$ is a function of $\hat{P}_{\boldsymbol{s}}$ and it might be different for a different state type class. Draw $2^{N(R+\rho(\hat{P}_{\boldsymbol{s}}))}$ random vectors independently from the type class $T^*_U(\hat{P}_{\boldsymbol{s}})$ induced by $\hat{P}^*_{\boldsymbol{u}}$, according to uniform distribution, where $\rho(\cdot)$ is a general bin-depth function. Arrange the vectors in an array with $M = 2^{NR}$ columns and $2^{N\rho(\hat{P}_{\boldsymbol{s}})}$ rows. The code $\mathcal{C}$ is the union of all sub-codes, i.e., $\mathcal{C} = \bigcup_{\hat{P}_{\boldsymbol{s}}} \mathcal{C}(\hat{P}_{\boldsymbol{s}})$. Note that the number of these sub-codes is polynomial in N. In this work, we choose $\rho(\hat{P}_{\boldsymbol{s}}) = I^*_{US}(\hat{P}_{\boldsymbol{s}}) + \epsilon$, where

$$I^*_{US}(\hat{P}_{\boldsymbol{s}}) = \sum_{u,s} \hat{P}_{\boldsymbol{s}}(s)P^*_{\boldsymbol{u}|\boldsymbol{s}}(u|s) \log \frac{\hat{P}_{\boldsymbol{s}}(s)P^*_{\boldsymbol{u}|\boldsymbol{s}}(u|s)}{\hat{P}^*_{\boldsymbol{u}}(u)} \ ,$$

i.e., $I^*_{US}(\hat{P}_{\boldsymbol{s}})$ is the mutual information $I(U; S)$ induced by $\hat{P}_{\boldsymbol{s}}(S) \cdot \hat{P}^*_{\boldsymbol{u}|\boldsymbol{s}}(U|S)$, and ϵ is an arbitrarily small positive constant. This choice ensures that the probability of encoding error vanishes at a double-exponentially rate [8, p. 1338].

The encoding of message m given a state sequence $\boldsymbol{s}$ is done in two steps: (i) Find an index l such that $\boldsymbol{u}_{l,m} \in \mathcal{C}(\hat{P}_{\boldsymbol{s}})$ is a member of the conditional type class $T^*_{\boldsymbol{u}_{l,m}|\boldsymbol{s}} = \{\boldsymbol{u}' : \hat{P}_{\boldsymbol{u}'\boldsymbol{s}} = \hat{P}^*_{\boldsymbol{u}_{l,m}|\boldsymbol{s}}\hat{P}_{\boldsymbol{s}}\}$. If more than one such l exists, pick one at random under the uniform distribution. If no such l can be found, pick $\boldsymbol{u}$ at random from $T^*_{\boldsymbol{u}|\boldsymbol{s}}$ under the uniform distribution. (ii) Draw $\boldsymbol{X}$ uniformly from $T^*_{\boldsymbol{x}|\boldsymbol{u}_{l,m}\boldsymbol{s}}$, induced by $\hat{P}^*_{\boldsymbol{xu}|\boldsymbol{s}}$ and $(\boldsymbol{u}_{l,m}, \boldsymbol{s})$. For notational simplicity, we use the shorthand λ to denote the type of state sequences $\hat{P}_{\boldsymbol{s}}$, and $\boldsymbol{u}_{\lambda,l,m}$ to denote $\boldsymbol{u}_{l,m} \in \mathcal{C}(\lambda)$.

In [8], a maximum penalized mutual information (MPMI) decoder was used to decode the above code. A MPMI decoder seeks a vector $\boldsymbol{u} \in \mathcal{C}$ that maximizes the penalized empirical mutual information criterion $\max_{\hat{P}_{\boldsymbol{S}}} \max_{\boldsymbol{u} \in \mathcal{C}(\hat{P}_{\boldsymbol{S}})} \left[I(\boldsymbol{u}; \boldsymbol{y}) - \psi(\hat{P}_{\boldsymbol{S}}) \right]$, where $\psi(\cdot)$ is a general penalty function. It was shown that the optimal choice of these functions is $\rho(\hat{P}_{\boldsymbol{S}}) = \psi(\hat{P}_{\boldsymbol{S}}) = I_{US}^*(\hat{P}_{\boldsymbol{S}}) + \epsilon$ where ϵ is an arbitrarily small positive constant. In this work, we assume that $\psi(\hat{P}_{\boldsymbol{S}}) = I_{US}^*(\hat{P}_{\boldsymbol{S}}) = I_{\hat{P}_{\boldsymbol{S}} P_{\boldsymbol{u}|\boldsymbol{S}}^*}^* (U; S)$ for reasons that will be given later. To allow decoding with an erasure/list option, we propose to modify the MPMI decoding rule in the spirit of (2). We choose $\rho(\hat{P}_{\boldsymbol{S}}) = \psi(\hat{P}_{\boldsymbol{S}}) = I_{US}^*(\hat{P}_{\boldsymbol{S}}) + \epsilon$ for reasons that will be discussed in Section 5.

2.2 The proposed decoding rule

For a given code $\mathcal{C}$ constructed as described in Subsection 2.1, we propose to use the following decoder $\varphi : \mathcal{Y}^N \to \{0, 1, \ldots, M\}$ with an erasure option: Declare m if

$$I(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) - I_{US}^*(\lambda) > T + \alpha \left| I(\boldsymbol{u}_{\lambda',l',m'}; \boldsymbol{y}) - I_{US}^*(\lambda') \right|^+ \quad \forall m' \neq m, \lambda', l', \tag{3}$$

otherwise, declare 0 (i.e., "erasure"), where $\alpha \geq 1$ and $T \geq 0$ are arbitrary parameters.

Our first step is to show that this decoder is unambiguous, i.e., at most one message index taken from $\{1, \ldots, M\}$ fulfills (3). This property is essential to allow decoding with an erasure option. This property is stated in the following Lemma:

Lemma 1. *For $\alpha \geq 1$ and $T \geq 0$, the proposed decoding rule* (3) *is unambiguous.*

The proof of Lemma 1 is deferred to the Appendix. Using a similar proof of Lemma 1, it can be shown that Csiszár and Körner's decoder (2) might be ambiguous if $0 < \lambda < 1$, contrary to the statement made in [2, Th. 5.11].

3 Erasure Option

Given a code $\mathcal{C}$, a decoder with an erasure option is a partition of $\mathcal{Y}^N$ into $(M+1)$ regions $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_M$. The decoder decides in favor of message m if $\boldsymbol{y} \in \mathcal{R}_m$, $m = 1, \ldots, M$, or it declares "erasure" if $\boldsymbol{y} \in \mathcal{R}_0$. Following Forney [1], let us define two error events. The event $\mathcal{E}_1$ is the event in which $\boldsymbol{y}$ does not fall in the decision region of the correct message. The event $\mathcal{E}_2$ is the event of *undetected error*, namely, the event in which $\boldsymbol{y}$ falls in $\mathcal{R}_{m'}$,

$m' \neq 0$, while m was transmitted. The probabilities of these error events are given by

$$\Pr\{\mathcal{E}_1\} = \frac{1}{M} \sum_{m=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{R}_m^c} P(\boldsymbol{y}|\boldsymbol{x}_m) \tag{4}$$

$$\Pr\{\mathcal{E}_2\} = \frac{1}{M} \sum_{m=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{R}_m} \sum_{m' \neq m} P(\boldsymbol{y}|\boldsymbol{x}_{m'}) \tag{5}$$

where $P(\boldsymbol{y}|\boldsymbol{x}_m) = \sum_{\boldsymbol{s} \in \mathcal{S}^N} P_S(\boldsymbol{s}) W(\boldsymbol{y}|\boldsymbol{x}_m(\boldsymbol{s}), \boldsymbol{s})$.

Let $J(P_S P_{UX|S} P_{Y|XS}) \triangleq I(U;Y) - I(U;S)$ where P_S, $P_{UX|S}$, and $P_{Y|XS}$ are three (conditional) probability distributions of the quadruplet RVs (U, S, X, Y). The following theorem presents exponential bounds on $\Pr\{\mathcal{E}_1\}$ and $\Pr\{\mathcal{E}_2\}$ for decoding with erasure option:

Theorem 1. *For every $\alpha \geq 1$ and $T \geq 0$ there exists a $N-$length block code of rate R such that the following error exponents can be achieved simultaneously*

$$\Pr\{\mathcal{E}_1\} \leq \exp_2\{-NE_1(R, W, T, \alpha)\} \tag{6}$$

$$\Pr\{\mathcal{E}_2\} \leq \exp_2\{-NE_2(R, W, T, \alpha)\} \tag{7}$$

where

$$E_1(R, W, T, \alpha) = \min_{\tilde{P}_S} \max_{\tilde{P}_{UX|S}} \min \left\{ \min_{\tilde{P}_{Y|XS}: J(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS}) \leq T} \mathcal{D}(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS} \| P_S \tilde{P}_{UX|S} W), \right.$$
$$\left. \min_{\tilde{P}_{Y|XS}} \left[\mathcal{D}(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS} \| P_S \tilde{P}_{UX|S} W) + \left| \frac{1}{\alpha} \left(J(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS}) - T \right) - R \right|^+ \right] \right\} \tag{8}$$

and

$$E_2(R, W, T, \alpha) = \min_{\tilde{P}_S} \max_{\tilde{P}_{UX|S}} \min_{\tilde{P}_{Y|XS}} \left\{ \right.$$
$$\left. \mathcal{D}(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS} \| P_S \tilde{P}_{UX|S} W) + \left| T + \alpha |J(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS})|^+ - R \right|^+ \right\}. \tag{9}$$

Proof. We analyze $\Pr\{\mathcal{E}_1\}$ and $\Pr\{\mathcal{E}_2\}$ using the proposed decoder (3). The proof is similar in some parts to the derivation done in the proof of Theorem 3.2 in [8], but it is given in full for the sake of completeness.

Fix a probability distribution $\hat{P}_{\boldsymbol{ux|s}}$ and construct a code $\mathcal{C}$ as described in Subsection 2.1. An encoding error occurs when the first encoding step fails. Namely, given m

and $\boldsymbol{s}$ there is no index l such that $\boldsymbol{u}_{\lambda,l,m} \in T^*_{\boldsymbol{u}|\boldsymbol{s}}$. Since $\boldsymbol{u}$ is drawn randomly according to uniform distribution from $T^*_U(\hat{P}_{\boldsymbol{s}})$ it follows that

$$\Pr\left\{\mathcal{E}_c(m)|\boldsymbol{s}\right\} = \left[1 - \Pr\left\{\boldsymbol{U} \in T^*_{\boldsymbol{u}|\boldsymbol{s}}|\boldsymbol{U} \in T^*_U(\hat{P}_{\boldsymbol{s}})\right\}\right]^{2^{N\rho(\hat{P}_{\boldsymbol{s}})}} \tag{10}$$

where $\mathcal{E}_c(m)$ denotes encoding error when message index m is encoded, and

$$\begin{aligned}\Pr\left\{\boldsymbol{U} \in T^*_{\boldsymbol{u}|\boldsymbol{s}}|\boldsymbol{U} \in T^*_U(\hat{P}_{\boldsymbol{s}})\right\} &= \frac{|T^*_{\boldsymbol{u}|\boldsymbol{s}}|}{T^*_U(\hat{P}_{\boldsymbol{s}})} \\ &\doteq 2^{-NI^*_{US}(\hat{P}_{\boldsymbol{s}})} .\end{aligned} \tag{11}$$

Since $\rho(\hat{P}_{\boldsymbol{s}})$ was chosen to be greater than $I^*_{US}(\hat{P}_{\boldsymbol{s}})$ by ϵ we get that probability of encoding error of message m given $\boldsymbol{s}$ is upper bounded by

$$\Pr\left\{\mathcal{E}_c(m)|T(\hat{P}_{\boldsymbol{s}})\right\} \leq \exp\left\{-2^{N\epsilon}\right\}, \tag{12}$$

namely, the probability of encoding error decays in a double-exponential rate. See step 1 in [8, p. 1338] for more details.

The undetected error probability can be expressed as follows

$$\begin{aligned}\Pr\{\mathcal{E}_2\} &= \frac{1}{M}\sum_{m=1}^{M}\Pr\{\mathcal{E}_2|m \text{ is to be sent}\} \\ &= \Pr\left\{\mathcal{E}_2(1)\big|\, m=1\right\} \\ &= \sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}}|\, m=1)\Pr\left\{\mathcal{E}_2(1)|T_{\boldsymbol{usxy}}, m=1\right\} \\ &\leq \sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}}|\, m=1)\left[\Pr\left\{\mathcal{E}_c(1)|T_{\boldsymbol{usxy}}, m=1\right\}\right. \\ &\quad\left. + \Pr\left\{\mathcal{E}_2(1)|T_{\boldsymbol{usxy}}, m=1, \mathcal{E}_c(1)^c\right\}\right], \end{aligned} \tag{13}$$

where $\mathcal{E}_2(1)$ is the event of undetected error given that $m=1$ was sent, $T_{\boldsymbol{usxy}}$ is the joint typical class of the quadruplet $(\boldsymbol{u}, \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. Since all messages are drawn according the same probability distributions, the probability of a type class $T_{\boldsymbol{usxy}}$ is independent of the message index $1 \leq m \leq M$. Therefore,

$$\begin{aligned}P(T_{\boldsymbol{usxy}}|\, m=1) &= \exp_2 P(T_{\boldsymbol{usxy}}) \\ &\doteq \exp_2\{-N\mathcal{D}(\hat{P}_{\boldsymbol{s}}\hat{P}_{\boldsymbol{ux}|\boldsymbol{s}}\hat{P}_{\boldsymbol{y}|\boldsymbol{xs}}\|P_S\hat{P}_{\boldsymbol{ux}|\boldsymbol{s}}W)\} , \end{aligned} \tag{14}$$

as was shown in [8, eq.(5.12)]. An undetected error *can* occur only if there is a $\boldsymbol{u}_{\lambda',l',m'} \in \mathcal{C}$ such that

$$I(\boldsymbol{u}_{\lambda',l',m'}; \boldsymbol{y}) - I_{US}^*(\lambda') > T + \alpha |I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}})|^+ \, , \tag{15}$$

conditioned on $\boldsymbol{u}, \boldsymbol{s}, \boldsymbol{y}$ and $T_{\boldsymbol{usxy}}$.

Following [8, eq. (5.13)], the undetected error probability is upper bounded by

$$\Pr\left\{ \mathcal{E}_2(1) \big| T_{\boldsymbol{usxy}}, m = 1, \mathcal{E}_c(1)^c \right\} = 1 - \prod_{\hat{P}_{\boldsymbol{s}'}} \left[1 - P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}}) \right]^{2^{NI_{US}^*(\hat{P}_{\boldsymbol{s}'})}(2^{NR}-1)} ,\tag{16}$$

where $P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})$ is the probability that for some l' and $m' \neq 1$, $\boldsymbol{u}_{l',m'} \in \mathcal{C}(\hat{P}_{\boldsymbol{s}'})$ fulfills (15) conditioned on $\boldsymbol{u}, \boldsymbol{y}$ and $T_{\boldsymbol{usxy}}$. $P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})$ can be expressed as follows

$$
\begin{aligned}
P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}}) &= \sum_{\boldsymbol{u}' \in \mathcal{U}_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} P(\boldsymbol{u}'|\hat{P}_{\boldsymbol{s}'}) \\
&= \sum_{\boldsymbol{u}' \in \mathcal{U}_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} |T_U^*(\hat{P}_{\boldsymbol{s}'})|^{-1}
\end{aligned}
$$

where

$$\mathcal{U}_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}}) = \left\{ \boldsymbol{u}' \in T_U^*(\hat{P}_{\boldsymbol{s}'}) \, : \, I(\boldsymbol{u}'; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}'}) > T + \alpha \big| I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}}) \big|^+ \right\} .$$

$$\tag{17}$$

Clearly, $\mathcal{U}_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})$ is contained in the following set of conditional types

$$
\begin{aligned}
\mathcal{T}_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}}) &= \Big\{ T_{\boldsymbol{u}'|\boldsymbol{y}} : T_{\boldsymbol{u}'} = T_U^*(\hat{P}_{\boldsymbol{s}'}), \\
&\qquad I(\boldsymbol{u}'; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}'}) > T + \alpha \big| I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}}) \big|^+ \Big\} \\
&\subseteq \Big\{ T_{\boldsymbol{u}'|\boldsymbol{y}} : I(\boldsymbol{u}'; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}'}) > T + \alpha \big| I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}}) \big|^+ \Big\} .
\end{aligned}
$$

Using similar steps as in [8, eq. (5.14)-(5.17)], we get

$$
\begin{aligned}
&P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}}) \\
&\leq \sum_{T_{\boldsymbol{u}'|\boldsymbol{y}} \subseteq \mathcal{T}_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} \frac{|T_{\boldsymbol{u}'|\boldsymbol{y}}|}{|T_{\boldsymbol{u}'}|} \\
&\doteq \sum_{T_{\boldsymbol{u}'|\boldsymbol{y}} \subseteq \mathcal{T}_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} 2^{-NI(\boldsymbol{u}'; \boldsymbol{y})} \\
&\leq \exp_2\left\{ -N\Big[T + \alpha \big| I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}}) \big|^+ + I_{US}^*(\hat{P}_{\boldsymbol{s}'}) \Big] \right\} . \tag{18}
\end{aligned}
$$

Applying the following bound [8, eq. (5.18)], which can be regarded as a generalized union bound,

$$1 - \prod_i (1 - \alpha_i)^{t_i} \leq \min\left\{1, \sum_i \alpha_i t_i\right\}, \qquad 0 \leq \alpha_i \leq 1, t_i \geq 1 \tag{19}$$

on Eq.(18) we get

$$
\begin{aligned}
\Pr\left\{\mathcal{E}_2(1)|T_{\boldsymbol{usxy}}, \mathcal{E}_c(1)^c\right\} &= 1 - \prod_{\hat{P}_{\boldsymbol{s}'}}\left[1 - P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})\right]^{2^{NI^*_{US}(\hat{P}_{\boldsymbol{s}'})}(2^{NR}-1)} \\
&\leq \min\left\{1, \sum_{\hat{P}_{\boldsymbol{s}'}} P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, T_{\boldsymbol{usxy}})2^{NI^*_{US}(\hat{P}_{\boldsymbol{s}'})}(2^{NR}-1)\right\} \\
&\leq \exp_2\left\{-N\left|T + \alpha|I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}})|^+ - R\right|^+\right\}. \tag{20}
\end{aligned}
$$

Combining (13), (14), (20) and optimizing over $\hat{P}_{\boldsymbol{ux}|\boldsymbol{s}}$ and $\hat{P}_{\boldsymbol{s}}$ we get that

$$\Pr\{\mathcal{E}_2\} \leq \exp_2\left\{-NE_2(R, W, T, \alpha)\right\}, \tag{21}$$

where $\mathcal{E}_2(R, W, T, \alpha)$ is given in (9).

Similarly to derivation of $E_2(R, W, T, \alpha)$, we can upper bound the probability of not making the right decision, denoted by $\Pr\{\mathcal{E}_1\}$. This error event occurs when the received $\boldsymbol{y}$ does not belong to the decision region corresponding to the transmitted message m. Therefore, an error occurs when

$$I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) \leq T + \alpha|I(\boldsymbol{u}_{\lambda', l', m'}; \boldsymbol{y}) - I^*_{US}(\lambda')|^+ \tag{22}$$

for some λ', l' and $m' \neq m$, conditioned on $\boldsymbol{u}, \boldsymbol{s}, \boldsymbol{y}$ and $T_{\boldsymbol{usxy}}$. This happens if and only if

$$I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\lambda) \leq T, \tag{23}$$

or

$$T < I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\lambda) \leq T + \alpha|I(\boldsymbol{u}_{\lambda', l', m'}, \boldsymbol{y}) - I^*_{US}(\lambda')|^+. \tag{24}$$

Following (24), $I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\lambda) - T$ is strictly positive since $T \geq 0$ and $I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\lambda)$ is strictly greater than T. Moreover, (24) implies that $I(\boldsymbol{u}_{\lambda', l', m'}, \boldsymbol{y}) - I^*_{US}(\lambda')$ must be positive too since

$$T < T + \alpha|I(\boldsymbol{u}_{\lambda', l', m'}, \boldsymbol{y}) - I^*_{US}(\lambda')|^+ \quad \Rightarrow \quad \alpha|I(\boldsymbol{u}_{\lambda', l', m'}, \boldsymbol{y}) - I^*_{US}(\lambda')|^+ > 0, \ a \geq 1$$

which means that the clipping function $|\cdot|^+$ was not active, namely, $I(\boldsymbol{u}_{\lambda',l',m'}, \boldsymbol{y}) - I^*_{US}(\lambda') > 0$. Therefore (24) implies that

$$\frac{1}{\alpha}\Big[I(\boldsymbol{u};\boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) - T\Big] \le I(\boldsymbol{u}_{\lambda',l',m'};\boldsymbol{y}) - I^*_{US}(\lambda') \,. \tag{25}$$

Hence, the event of not making the right decision is a union of two disjoint events (23) and (25). Therefore,

$$
\begin{aligned}
\Pr\{\mathcal{E}_1\} &= \frac{1}{M}\sum_{m=1}^M \Pr\Big\{\mathcal{E}_1\big|m \text{ is to be sent}\Big\} \\
&= \Pr\Big\{\mathcal{E}_1(1)\big|m=1\Big\} \\
&= \sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}})\Pr\Big\{\mathcal{E}_1(1)\big|T_{\boldsymbol{usxy}},m=1\Big\} \\
&\le \sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}})\Big[\Pr\Big\{\mathcal{E}_c(1)\big|T_{\boldsymbol{usxy}},m=1\Big\} + \mathbb{1}\big\{I(\boldsymbol{u};\boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) \le T\big\} \\
&\qquad + \Pr\Big\{\mathcal{A}(1)\big|T_{\boldsymbol{usxy}},m=1,\mathcal{E}_c(1)^c\Big\}\Big] \,,
\end{aligned}
\tag{26}
$$

where $\mathcal{E}_1(1)$ is the event of making the wrong decision given that $m=1$, and $\mathcal{A}(1)$ is the event in which $T < I(\boldsymbol{u};\boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) \le T + \alpha[I(\boldsymbol{u}_{\lambda',l',m'};\boldsymbol{y}) - I^*_{US}(\lambda')]$ for some λ', l' and $m' \ne 1$. The last sum can be rewritten as follows

$$
\begin{aligned}
\sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}}|\, m=1)&\Big[\Pr\Big\{\mathcal{E}_c(1)\big|T_{\boldsymbol{usxy}},m=1\Big\} + \Pr\Big\{\mathcal{A}(1)\big|T_{\boldsymbol{usxy}},m=1,\mathcal{E}_c(1)^c\Big\}\Big] \\
&+ \sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}}|\, m=1)\mathbb{1}\big\{I(\boldsymbol{u};\boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) \le T\big\} \,.
\end{aligned}
$$
$$\tag{27}$$

The second summand of (27) can easily be estimated using the method of types and by applying (14)

$$
\begin{aligned}
\sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}}|\, m=1)&\mathbb{1}\big\{I(\boldsymbol{u};\boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) \le T\big\} = \\
&= \sum_{T_{\boldsymbol{usxy}}:I(\boldsymbol{u};\boldsymbol{y})-I^*_{US}(\hat{P}_{\boldsymbol{s}})\le T} P(T_{\boldsymbol{usxy}}) \\
&\doteq \exp_2\Big\{-N \min_{\hat{P}_{\boldsymbol{usxy}}:I(\boldsymbol{u};\boldsymbol{y})-I^*_{US}(\hat{P}_{\boldsymbol{s}})\le T} \mathcal{D}(\hat{P}_{\boldsymbol{s}}\hat{P}_{\boldsymbol{ux}|\boldsymbol{s}}\hat{P}_{\boldsymbol{y}|\boldsymbol{xs}}\|P_S\hat{P}_{\boldsymbol{ux}|\boldsymbol{s}}W)\Big\} \tag{28}
\end{aligned}
$$

As for the first summand of (26), it can be upper bounded similarly to the undetected error probability in the following way

$$\Pr\left\{\mathcal{A}(1)\big|T_{\boldsymbol{usxy}}, m=1, \mathcal{E}_c(1)^c\right\} = 1 - \prod_{\hat{P}_{\boldsymbol{s}'}}\left[1 - P_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})\right]^{2^{NI^*_{US}(\hat{P}_{\boldsymbol{s}'})}(2^{NR}-1)} \quad (29)$$

where $P_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})$ is the probability that $\boldsymbol{u}_{l',m'} \in \mathcal{C}(\hat{P}_{\boldsymbol{s}'})$ fulfills (25) conditioned on $\boldsymbol{u}, \boldsymbol{y}$ and $T_{\boldsymbol{usxy}}$ for some l' and $m' \neq 1$. Therefore,

$$\begin{aligned}
P_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}}) &= \sum_{\boldsymbol{u}' \in \mathcal{U}_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} P(\boldsymbol{u}'|\hat{P}_{\boldsymbol{s}'}) \\
&= \sum_{\boldsymbol{u}' \in \mathcal{U}_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} |T^*_U(\hat{P}_{\boldsymbol{s}'})|^{-1} \quad (30)
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{U}_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}}) = \\
\left\{\boldsymbol{u}' \in T^*_U(\hat{P}_{\boldsymbol{s}'}) : I(\boldsymbol{u}'; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}'}) > \frac{1}{\alpha}[I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) - T]\right\}. \quad (31)
\end{aligned}$$

Again, $\mathcal{U}_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})$ is contained in the following set of conditional types

$$\begin{aligned}
\mathcal{T}_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}}) &= \left\{T_{\boldsymbol{u}'|\boldsymbol{y}} : T_{\boldsymbol{u}'} = T^*_U(\hat{P}_{\boldsymbol{s}'}),\right. \\
&\qquad \left. I(\boldsymbol{u}'; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}'}) > \frac{1}{\alpha}[I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) - T]\right\} \\
&\subseteq \left\{T_{\boldsymbol{u}'|\boldsymbol{y}} : I(\boldsymbol{u}'; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}'}) > \frac{1}{\alpha}[I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) - T]\right\}.
\end{aligned}$$

Using similar steps as in the first part of the proof, we get that

$$\begin{aligned}
&P_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}}) \\
&\leq \sum_{T_{\boldsymbol{u}'|\boldsymbol{y}} \subseteq \mathcal{T}_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} \frac{|T_{\boldsymbol{u}'|\boldsymbol{y}}|}{|T_{\boldsymbol{u}'}|} \\
&\doteq \sum_{T_{\boldsymbol{u}'|\boldsymbol{y}} \subseteq \mathcal{T}_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, \hat{P}_{\boldsymbol{usxy}})} 2^{-NI(\boldsymbol{u}';\boldsymbol{y})} \\
&\leq \exp_2\left\{-N\left[\frac{1}{\alpha}\big(I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) - T\big) + I^*_{US}(\hat{P}_{\boldsymbol{s}'})\right]\right\}. \quad (32)
\end{aligned}$$

Applying the union bound (19), we get

$$\begin{aligned}
\Pr\left\{\mathcal{E}_1(1)|T_{\boldsymbol{usxy}}, \mathcal{E}_c(1)^c\right\} &= 1 - \prod_{\hat{P}_{\boldsymbol{s}'}}\left[1 - P_{e_1}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})\right]^{2^{NI^*_{US}(\hat{P}_{\boldsymbol{s}'})}(2^{NR}-1)} \\
&\leq \exp_2\left\{-N\left|\frac{1}{\alpha}[I(\boldsymbol{u}; \boldsymbol{y}) - I^*_{US}(\hat{P}_{\boldsymbol{s}}) - T] - R\right|^+\right\}. \quad (33)
\end{aligned}$$

Combining (26), (28), (14),(33), and optimizing over $\hat{P}_{\boldsymbol{ux}|\boldsymbol{s}}$ and $\hat{P}_{\boldsymbol{s}}$ we get that

$$\Pr\{\mathcal{E}_1\} \leq \exp_2\left\{-NE_1(R, W, T, \alpha)\right\} , \tag{34}$$

where $E_1(R, W, T, \alpha)$ is given in (8).

$\square$

4 List Decoding

A decoder with a variable size list produces a list of candidate estimates for the correct message. Let $\mathcal{E}_1$ denote the list error event, namely, the event in which $\boldsymbol{y}$ does not fall in the decision region corresponding to the correct message. As stated by Forney [1, p. 206], the event corresponding to $\mathcal{E}_2$ under decoding with an erasure option, is the average number of incorrect messages on the list, denoted by $\bar{N}_I$, where

$$\bar{N}_I = \sum_{m'} \Pr\{m' \text{ is on the list and incorrect}\} . \tag{35}$$

In the following Theorem, exponential bounds are offered for the probability of $\mathcal{E}_1$ and on the average number of incorrect messages on the list. These bounds are obtained by generalizing the proposed decoding rule (3) to the variable list size case by extending the range of its parameters α and T.

Theorem 2. *For every $\alpha \in (0, 1)$ and $T \in \mathbb{R}$ there exists a $N-$length block code of rate R such that the following error exponents can be achieved simultaneously*

$$\Pr\{\mathcal{E}_1\} \leq \exp_2\left\{-NE_1(R, W, T, \alpha)\right\} \tag{36}$$

$$\bar{N}_I \leq \exp_2\left\{-NE_2(R, W, T, \alpha)\right\} , \tag{37}$$

where

$$E_1(R, W, T, \alpha) = \min_{\tilde{P}_S} \max_{\tilde{P}_{UX|S}} \min \left\{ \min_{\tilde{P}_{Y|XS}: J(\tilde{P}_S\tilde{P}_{UX|S}\tilde{P}_{Y|XS}) \leq T} \mathcal{D}(\tilde{P}_S\tilde{P}_{UX|S}\hat{P}_{Y|XS}\|P_S\tilde{P}_{UX|S}W), \right.$$

$$\left. \min_{\tilde{P}_{Y|XS}} \left(\mathcal{D}(\tilde{P}_{Y|XS}\|W|\tilde{P}_S\tilde{P}_{UX|S}) + \left|\frac{1}{\alpha}[J(\tilde{P}_S\tilde{P}_{UX|S}\tilde{P}_{Y|XS}) - T] - R\right|^+\right) \right\} , \tag{38}$$

and

$$E_2(R, W, T, \alpha) = \min_{\tilde{P}_S} \max_{\tilde{P}_{UX|S}} \min_{\tilde{P}_{Y|XS}} \left\{ \mathcal{D}\big(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS} \| P_S \tilde{P}_{UX|S} W\big) \right.$$

$$\left. + \big|T + \alpha |J(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS})|^+\big|^+ - R \right\}. \quad (39)$$

Proof. To allow list option we take $\alpha \in (0, 1)$ and $T \in \mathbb{R}$. Therefore, the following decoding rule will be used: add m to the list if

$$I(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) - I_{US}^*(\lambda) > T + \alpha |I(\boldsymbol{u}_{\lambda',l',m'}; \boldsymbol{y}) - I_{US}^*(\lambda')|^+$$

for all $\lambda', l', m' \neq m$. An empty list is regarded as "erasure".

Fix a probability distribution $\hat{P}_{\boldsymbol{ux}|\boldsymbol{s}}$ and construct a code $\mathcal{C}$ as described in Subsection 2.1. The encoding error is described in (12). Let us start with the probability of list error. A list error occurs when

$$I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}}) \leq T + \alpha \big|I(\boldsymbol{u}_{\lambda',l',m'}; \boldsymbol{y}) - I_{US}^*(\lambda')\big|^+$$

for some l', λ' and $m' \neq m$, conditioned on $\boldsymbol{u}, \boldsymbol{s}, \boldsymbol{y}$ and $T_{\boldsymbol{usxy}}$. This happens if

$$I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\lambda) \leq T$$

or

$$T < I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\lambda) \leq T + \alpha[I(\boldsymbol{u}_{\lambda',l',m'}, \boldsymbol{y}) - I_{US}^*(\lambda')] \,,$$

similarly to the second part of the proof of Theorem 1. From this point on we follow the derivation of $\Pr\{\mathcal{E}_1\}$ in the proof of Theorem 1, and obtain the desired exponent (38).

Our next step is to upper bound the average number of incorrect words on the list $\bar{N}_I$. Following [1, eq. (12)-(13)], the average number of incorrect codewords is

$$
\begin{aligned}
\bar{N}_I &= \frac{1}{M} \sum_{m=1}^{M} \sum_{m' \neq m} \Pr\left\{ m' \text{ is on list} \,\big|\, m \text{ was sent} \right\} \\
&= \sum_{m' > 1} \Pr\left\{ m' \text{ is on list} \,\big|\, m = 1 \right\} \\
&= (M - 1) \sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}} | m = 1) \Pr\left\{ m' \neq m \text{ is on list} \,\big|\, m = 1, T_{\boldsymbol{usxy}} \right\}
\end{aligned}
$$

$$(40)$$

where the second equality is because the messages are equiprobable. The probability that $m' > 1$ is on the list given that $m = 1$ was sent can be bounded as follows

$$\Pr\left\{m' \text{ is on list}\big|\, m = 1\right\} \leq \Pr\left\{\mathcal{E}_c(1)\big|T_{\boldsymbol{usxy}}, m = 1\right\}$$
$$+ \Pr\left\{m' \neq m \text{ is on list}\big|T_{\boldsymbol{usxy}}, m = 1, \mathcal{E}_c(1)^c\right\} \tag{41}$$

where $\mathcal{E}_c(1)$ is the encoding error event of message $m = 1$. Applying (14) we get

$$\bar{N}_I \;\; \leq \;\; M \sum_{T_{\boldsymbol{usxy}}} P(T_{\boldsymbol{usxy}}) \Big[\Pr\left\{\mathcal{E}_c(1)\big|T_{\boldsymbol{usxy}}, m = 1\right\}$$
$$+ \Pr\left\{m' \neq m \text{ is on list}\big|T_{\boldsymbol{usxy}}, m = 1, \mathcal{E}_c(1)^c\right\} \Big] . \tag{42}$$

The probability that $m' \neq 1$ is on the decoding list given that $m = 1$ was sent successfully and given $T_{\boldsymbol{usxy}}$ is upper bounded by

$$\Pr\left\{m' \neq m \text{ is on list}\big|T_{\boldsymbol{usxy}}, m = 1, \mathcal{E}_c(1)^c\right\} =$$
$$1 - \prod_{\hat{P}_{\boldsymbol{s}'}} \left[1 - P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})\right]^{2^{NI_{US}^*(\hat{P}_{\boldsymbol{s}'})}} , \quad (43)$$

where $P_{e_2}(\boldsymbol{u}, \boldsymbol{y}, \hat{P}_{\boldsymbol{s}'}, T_{\boldsymbol{usxy}})$ is the probability that there exist $\boldsymbol{u}_{l',m'} \in \mathcal{C}(\hat{P}_{\boldsymbol{s}'})$ which obeys the decoding rule conditioned on $\boldsymbol{u}, \boldsymbol{y}$ and $T_{\boldsymbol{usxy}}$ for some l'. Namely, there exist a codeword $\boldsymbol{u}_{l',m'}$ which "beats" all other codewords from a different column. This probability is upper bounded by the event in which $\boldsymbol{u}_{l',m'} \in \mathcal{C}(\hat{P}_{\boldsymbol{s}'})$ "beats" *only* the correct codeword, i.e.,

$$I(\boldsymbol{u}_{\lambda',l',m'}; \boldsymbol{y}) - I_{US}^*(\hat{P}_{\boldsymbol{s}'}) > T + \alpha \big|I(\boldsymbol{u}; \boldsymbol{y}) - I_{US}^*(\lambda)\big|^+ ,$$

which in turn upper bounds (43). Therefore,

From this point on we follow the derivation of $\Pr\{\mathcal{E}_2\}$ in the proof of Theorem 1. Note that M multiplies $\Pr\left\{m' \neq m \text{ is on list}\big|T_{\boldsymbol{usxy}}, m = 1, \mathcal{E}_c(1)^c\right\}$ in (40). Therefore, the coding rate R is found outside the clipping function $|\cdot|^+$ in (39) unlike (9). This implies that $\bar{N}_I$ might be greater than unity as expected.

$\square$

5   Discussion

In this paper, we proposed universally achievable error exponents for decoding with an erasure option and a variable size list. These results were obtained by examining a universal

decoder with an erasure option, inspired by Csiszár and Körner's [2, p. 176] for DMCs. By changing the decoder's parameters, one can switch from list decoding ($T \in \mathbb{R}, \alpha \in (0,1)$) to decoding with an erasure option ($T \geq 0, \alpha \geq 1$). A similar behavior was exemplified by Forney [1] with the optimal decoding rule for DMCs. The proposed decoder (3) has a similar structure to Csiszár and Körner's decoder (2), however, it does not depend on the coding rate R which make it more general.

Setting specific values to α and T achieves some known results. If we take $\alpha = 1$ and $T = 0$ in Theorem 1, we get that $E_1(R, W, T, \alpha) = E_2(R, W, T, \alpha) = E(R, W)$, where $E(R, W)$ is the exponent achieved in [8, Th. 3.2] for a known channel.

In [1, eq.(11a)], Forney proposed a suboptimal decoding rule with an erasure option in which the decoder declares "m" if

$$\Pr(\boldsymbol{y}, \boldsymbol{x}_m) \geq e^{NT} \Pr(\boldsymbol{y}, \boldsymbol{x}_{m_2}) \ , \tag{44}$$

otherwise, an "erasure" is declared, where $\Pr(\boldsymbol{y}, \boldsymbol{x}_{m_2})$ is the probability of the second most likely code word, and T is a positive parameter. Hence, the probability of the most likely code word must be at least e^{NT} times higher than the probability of any other code word given $\boldsymbol{y}$. If we think of $\max_{l,\lambda}[I(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) - I_{US}^*(\lambda)]$ as the normalized logarithm of the empirical generalized a-post priori probability of message m given $\boldsymbol{y}$, as stated in [8, p. 1332], then by setting $\alpha = 1$ and $T \geq 0$ in (3), we obtain an empirical version of Forney's suboptimal decoding rule (44).

Unlike [8], we fixed the penalty function ψ and the bin-depth function ρ beforehand. Clearly, taking $\rho(\hat{P}_{\boldsymbol{s}}) = I_{US}^*(\hat{P}_{\boldsymbol{s}})$ is an optimal choice since it is the lowest exponential rate which ensures a vanishing encoding error probability. Higher values of ρ increases the probability of decoding error (see [8, p. 1331]). However, it is not clear whether $\psi = \rho$ is an optimal choice in (3) (at least not when $\alpha = 1$). If, for example, we derive the exponent

$\mathcal{E}_1$ in Th. 1 with a general penalty function $\psi(\cdot)$ we get that

$$E_1(R, W, T, \alpha, \psi) = \max_{\psi} \min_{\tilde{P}_S} \max_{\tilde{P}_{UX|S}} \min \Big\{$$

$$\min_{\tilde{P}_{Y|XS}:I(Y;U)-\psi(\tilde{P}_S) \leq T} \mathcal{D}(\tilde{P}_S \tilde{P}_{UX|S} \tilde{P}_{Y|XS} \| P_S \tilde{P}_{UX|S} W),$$

$$\min_{\tilde{P}_{Y|XS}} \Big[\mathcal{D}(\tilde{P}_{Y|XS} \| W | \tilde{P}_S \tilde{P}_{UX|S})$$

$$+ \Big| \frac{1}{\alpha} \big(I(Y;U) - \psi(\tilde{P}_S) - T \big) + \min_{\hat{P}_S}[\psi(\hat{P}_S) - \hat{I}(U;S)] - R \Big|^+ \Big] \Big\},$$

where $\hat{I}(U;S)$ is the mutual information induced by $\hat{P}_S(S)\tilde{P}_{U|S}(U|S)$. As one can see, ψ is involved in many places of the above expression and therefore cannot be easily optimized. Moreover, the argument used in [8] to prove that $\psi = \rho$ is optimal cannot be applied here. This calls for further investigation.

We note that Forney's derivation cannot be easily applied to state dependent channels where side information is present at the transmitter. The main difficulty arises from the fact that the overall channel from $\boldsymbol{u}$ to $\boldsymbol{y}$ is not memoryless.

6 Appendix

Proof of Lemma 1. Denote $J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) \triangleq I(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) - I^*_{US}(\lambda)$. Suppose that the lemma is false. Therefore, there are two vectors $\boldsymbol{u}_{\lambda,l,m}$ and $\boldsymbol{u}_{\hat{\lambda},\hat{l},\hat{m}}$ such that:

$$J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) > T + \alpha |J(\boldsymbol{u}_{\lambda',l',m'}; \boldsymbol{y})|^+ \quad \forall m' \neq m, \forall \lambda', l' , \tag{45}$$

$$J(\boldsymbol{u}_{\hat{\lambda},\hat{l},\hat{m}}; \boldsymbol{y}) > T + \alpha |J(\boldsymbol{u}_{\lambda',l',m'}; \boldsymbol{y})|^+ \quad \forall m' \neq \hat{m}, \forall \lambda', l' . \tag{46}$$

Hence,

$$J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) > T + \alpha |J(\boldsymbol{u}_{\hat{\lambda},\hat{l},\hat{m}}; \boldsymbol{y})|^+ \tag{47}$$

$$J(\boldsymbol{u}_{\hat{\lambda},\hat{l},\hat{m}}; \boldsymbol{y}) > T + \alpha |J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y})|^+ \qquad . \tag{48}$$

From (47)-(48) it is clear that

$$T + \alpha \big[J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) \big] < J(\boldsymbol{u}_{\hat{\lambda},\hat{l},\hat{m}}; \boldsymbol{y}) < \frac{1}{\alpha} \big[J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) - T \big] , \tag{49}$$

which implies that

$$\alpha T + \alpha^2 J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) < J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) - T$$

$$0 < (1 - \alpha^2) J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y}) - T(1 + \alpha) . \tag{50}$$

Clearly, the right hand side of (50) cannot be positive since $1 - \alpha^2 \leq 0$, $T(1 + \alpha) \geq 0$ and $J(\boldsymbol{u}_{\lambda,l,m}; \boldsymbol{y})$ is positive following (45) . Hence, the assumption that the decoding rule is ambiguous is wrong. Note that the same proof can be used to show that for $\lambda \geq 1$, Csiszár and Körner's decoder [2, Th. 5.11] is unambiguous. $\square$

References

[1] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.

[3] E. Telatar, "Multi-access communications with decision feedback decoding," Ph.D. dissertation, M.I.T., Cambridge, Massachusetts, May 1992.

[4] P. Moulin, "A Neyman-Pearson approach to universal erasure and list decoding," *submitted to IEEE Trans. on Information Theory*, Jan. 2008. [Online]. Available: http://arxiv.org/abs/0801.4544

[5] E. Telatar and R. G. Gallager, "New exponential upper bounds to error and erasure probabilities," in *Proc. of the International Symposium on Information Theory(ISIT '94)*, Trondheim, Norway, Jun. 1994, p. 379.

[6] E. Telatar, "Exponential bounds for list size moments and error probability," in *Proc. of the IEEE Information Theory Workshop*, Killarney, Ireland, Jun. 1998, p. 60.

[7] S. Gel'fand and M. Pinsker, "Coding for channels with random parameter," *Problems of Information and Control*, vol. 9, no. 1, pp. 19–31, 1980.

[8] P. Moulin and Y. Wang, "Capacity and random-coding error exponents for channel coding with side information," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1326–1347, Apr. 2007.