Error Exponents of Optimum Erasure/List and Ordinary Decoding for Channels with Side Information

Erez Sabbag
Densbits Technologies Ltd.
P.O. Box 15111,
Haifa 31015, Israel
Email: sabbage@gmail.com

Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
Email: merhav@ee.technion.ac.il

Abstract—In this work we analyze achievable random coding error exponents pertaining to erasure/list decoding for channels with side information present at the transmitter. The analysis is carried out using the *optimal* decoding rule proposed by Forney. A key ingredient in the analysis is the evaluation of moments of a certain distance enumerator. This approach leads to a new exponentially tight bounds. These results are obtained by exploring a random binning code with conditionally constant composition codewords previously proposed by Moulin and Wang. Later, these results are used to obtain an achievable random coding error exponent for ordinary decoding.

I. INTRODUCTION

Decoding with an erasure option and variable list size decoding are generalizations of the ordinary decoding rule in which the decoder gives one estimate for the transmitted message. 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 analyzed random coding error exponents of erasure/list decoding for discrete memoryless channels (DMC's). These bounds were obtained by exploring the optimal decoding rule [1, eq. (11)]

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

where $\Pr(\boldsymbol{y}, \boldsymbol{x}_m)$ is the joint probability of the channel output $\boldsymbol{y}$ and the codeword $\boldsymbol{x}_m$, $\mathcal{R}_m$ is the decision region of message m, and T is an arbitrary parameter. The bounds were obtained using Gallager's bounding techniques. It was shown that the erasure option and the list option are "two sides of the same coin", namely, by changing the value of T one can switch from list decoding ($T < 0$) to decoding with an erasure option ($T > 0$).

Later, Csiszár and Körner [2, p.175, Th. 5.11] derived for DMCs universally achievable error exponents for a decoder

with an erasure option. These bounds were obtained by exploring a generalized maximum mutual information (MMI) decoder [2, p. 164] which has an erasure option.

Recently, upper bounds on the error probabilities under generalized decoding rules were provided for linear block codes over memoryless symmetric channels in [3].

In a recent paper [4], a decoder with an erasure option and a variable size list decoder for channels with non-casual side information at the transmitter were considered. A universally achievable region of error exponents was presented for decoding with an erasure option using a parameterized decoder in the spirit of Csiszár and Körner's decoder. Moreover, the proposed decoding rule was generalized by extending the range of its parameters to allow variable size list decoding. This extension gives a unified treatment for erasure/list decoding. An achievable region of exponential bounds on the probability of list error and the average number of incorrect messages on the list were given. The results were obtained using a random binning code with conditionally constant composition codewords proposed by Moulin and Wang [5] with the proposed parameterized decoder.

While these results extended the work of Csiszár and Körner to channels with side information present at the transmitter non-causally and generalized it to the case of variable list size decoding, the results in [4] were based on heuristic decoding rule with no proven relation to the optimal decoding rule. Moreover, while in [1] both error exponents were achieved by the same codebook distribution, in [4] the two error exponents might be achieved by different codebook distributions. This fact may indicate that the results of [4] are not tight. These two weak spots of [4] were the main motivation for the current work.

When reviewing Forney's proof [1, p.218] it is evident that the proof can be applied to the case of channels with side-information present at the transmitter if we could handle the analysis of two expressions: (i) $\mathbb{E}_{\mathcal{C}} P^\lambda(\boldsymbol{y}|m)$, namely, the moments of the probability of the channel output given that the correct message was sent, and (ii) $\mathbb{E}_{\mathcal{C}} \left(\sum_{m' \neq m} P(\boldsymbol{y}|m') \right)^\lambda$ the moments of the sum of the probabilities of all competing

messages, where the expectation is taken with respect to the code ensemble. Forney tackled these expressions using the Hölder's and Jensen's inequalities and the fact that the codewords were chosen in a symbol-by-symbol fashion. However, when the binning technique is introduced to the encoding rule $\boldsymbol{x} = f(m, \boldsymbol{s})$, which maps a channel state sequence $\boldsymbol{s}$ and a message index m into a channel input sequence $\boldsymbol{x}$ using an auxiliary sequence $\boldsymbol{u}$, Forney's analysis does not seem to be applicable.

To overcome these difficulties a different approach was taken. In this work we apply the distance enumerator approach (see [6] for detailed tutorial on the subject). This approach was applied in recent years to other problems in Information Theory with considerable success (see [7], [8], [9], [10]). In [11], the analysis of random coding error exponents for erausre/list decoding for DMCs was revisited, where the distance enumerator approach was used instead of the use of Hölder's and Jensen's inequalities as done by Forney. The resulting bounds, which are at least as tight as Forney's bounds, are simpler in the sense that it involves an optimization over one parameter only as opposed to Forney's bounds which involves two parameters. Moreover, when applying this technique to a certain universal decoder with erasures the tightness of these bounds is exemplified.

In this work we use the distance enumerator approach to analyze the optimal decoding rule proposed by Forney while adopting a random binning code with conditionally constant composition codewords (CCC) proposed by Moulin and Wang [5]. Later, these results are applied on Gallager's well-known upper bound presnted in [12] to get an achievable error exponent for the above channel model and code construction.

The outline of this work is the following: In Section II we present some notation which will be used throughout the paper, present the channel model, describe the codebook construction and the decoding rule which will be used during this paper. In Section III we present the main result of the paper. Section IV describes the case of ordinary decoding rule and presents the corresponding achievable error exponent. In Section V we discuss the result and present some directions for future research.

II. NOTATION AND PRELIMINARIES

We begin with some notation 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).

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

Let the vector $\hat{P}_{\boldsymbol{x}} = \left\{ \hat{P}_{\boldsymbol{x}}(a), \ a \in \mathcal{X} \right\}$ 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 $\mathcal{B}$, where $|\mathcal{B}| < \infty$ is defined as

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

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.

A. Channel Model and Code Construction

In this section, we describe the channel model and the code construction. Note that the channel model is the same channel model described in [4], and the code construction is almost identical. However, the decoding rule used in this work is based on the optimal decoding rule, while in [4] a sub-optimal decoding rule was used.

We consider a discrete memoryless state-dependent channel with a finite input set $\mathcal{X}$, a finite state alphabet $\mathcal{S}$, a finite output alphabet $\mathcal{Y}$, and a transition probability 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 $m \in \{1, \ldots, M\}$, where $M = \lfloor \exp\{NR\} \rfloor$ and R is the code rate, 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. We note that the channel is fixed and known to all parties, as presented in the work of Gel'fand and Pinsker [13].

In [5, p. 1337], Moulin and Wang used in their derivation a binning codebook with conditionally constant composition (CCC) codewords. A similar codebook will be used in our proofs. For the sake of completeness, we briefly describe the codebook construction and the encoding process. The decoding part will be described in detail later. The codebook construction requires the use of an auxiliary random variable $U \in \mathcal{U}$ which takes on values in a finite set of size $|\mathcal{X}||\mathcal{S}|$ as will be shown later.

For a given empirical conditional distribution $\hat{P}^*_{\boldsymbol{u}|\boldsymbol{s}}$, a sub-codebook $\mathcal{C}(\hat{P}_{\boldsymbol{s}})$ 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

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

where $\hat{P}_{\boldsymbol{s}}$ is the empirical distribution induced by $T_{\boldsymbol{s}}$. Note that $\hat{P}^*_{\boldsymbol{u}}(u)$ is a function of $\hat{P}_{\boldsymbol{s}}$ and it might be different for other state type classes. Draw $\exp\{N[R + \rho(\hat{P}_{\boldsymbol{s}})]\}$ random vectors independently from the type class $T^*_U(\hat{P}_{\boldsymbol{s}})$ induced by $\hat{P}^*_{\boldsymbol{u}}$, according to uniform distribution where $\rho(\hat{P}_{\boldsymbol{s}}) = I^*_{US}(\hat{P}_{\boldsymbol{s}}) + \epsilon$, and

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

This choice ensures that the probability of encoding error vanishes at a double-exponentially rate [5, p. 1338]. Arrange the vectors in an array with $M = \lfloor \exp\{NR\} \rfloor$ columns and $\lfloor \exp\{N\rho(\hat{P}_{\boldsymbol{s}})\} \rfloor$ rows. The codebook $\mathcal{C}$ is the union of all sub-codebooks, i.e., $\mathcal{C} = \bigcup_{\hat{P}_{\boldsymbol{s}}} \mathcal{C}(\hat{P}_{\boldsymbol{s}})$. Note that the number of these sub-codebooks is polynomial in N (at most $(N+1)^{|\mathcal{S}|}$).

To encode message m given a state sequence $\boldsymbol{s}$, the following two steps are done: (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) The channel input sequence is given by $\boldsymbol{x} = \big(x(s_1, u_1), x(s_2, u_2), \ldots, x(s_N, u_N)\big)$ where $x : \mathcal{S} \times \mathcal{U} \to \mathcal{X}$ is some fixed function. Since there is a deterministic mapping from $\mathcal{S} \times \mathcal{U}$ we can define a direct channel $\tilde{W} : \mathcal{S} \times \mathcal{U} \to \mathcal{Y}$ as

$$\tilde{W}(y|s,u) \triangleq W\big(y|s, x(s,u)\big) .$$

B. Optimal Decoding Rule

Given a codebook $\mathcal{C}$, 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, namely, the event in which the decoder decides wrongly. 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 $m \neq m'$ was transmitted. The average probabilities of these error events are given by

$$\Pr\{\mathcal{E}_1\} = \frac{1}{M} \sum_{m=1}^M \sum_{\boldsymbol{y} \in \mathcal{R}_m^c} P(\boldsymbol{y}|m) \qquad (2)$$

$$\Pr\{\mathcal{E}_2\} = \frac{1}{M} \sum_{m=1}^M \sum_{\boldsymbol{y} \in \mathcal{R}_m} \sum_{m' \neq m} P(\boldsymbol{y}|m) \qquad (3)$$

where $P(\boldsymbol{y}|m) = \sum_{\boldsymbol{s} \in \mathcal{S}^N} P_S(\boldsymbol{s})W\big(\boldsymbol{y}|\boldsymbol{x}_m(\boldsymbol{s}), \boldsymbol{s}\big)$, and the average probability of erasure is given by:

$$\Pr\{\mathcal{R}_0\} = \Pr\{\mathcal{E}_1\} - \Pr\{\mathcal{E}_2\} . \qquad (4)$$

In [1], Forney showed that the optimal tradeoff between the two error events is attained by the following decoding rule [1, p.208]:

$$\boldsymbol{y} \in \mathcal{R}_m^* \text{ iff } \frac{P(\boldsymbol{y}|m)}{\sum_{m' \neq m} P(\boldsymbol{y}|m')} \geq \exp\{NT\} , \qquad (5)$$

otherwise, declare "erasure" (i.e., $\boldsymbol{y} \in \mathcal{R}_0^*$), where $T \geq 0$ is a parameter which controls the trade-off between $\mathcal{E}_1$ and $\mathcal{E}_2$. When taking a closer look on the decoding rule derived by Forney [1, p.208], it is clear that these results can be applied to the channels with side-information available at the encoder non-causally as describe above where the only difference is that the probability of received channel output $\boldsymbol{y}$ given a message index m, denoted by $P(\boldsymbol{y}|m)$, is more involved.

For a given channel output sequence $\boldsymbol{y}$ and a message index m define

$$P(\boldsymbol{y}|m) = \sum_{\boldsymbol{s} \in \mathcal{S}^N} P_S(\boldsymbol{s})\tilde{W}\big(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{U}(\boldsymbol{s},m)\big) , \quad m \in \{1, \ldots, M\} \qquad (6)$$

where $\boldsymbol{U}(\boldsymbol{s}, m)$ is the encoding rule. Following the first steps of the derivation of [1, p.218], for a given codebook we get:

$$\Pr\{\mathcal{E}_1\} = \frac{1}{M} \sum_{m=1}^M \sum_{\boldsymbol{y} \in \mathcal{R}_m^c} P(\boldsymbol{y}|m) \qquad (7)$$

$$= \frac{1}{M} \sum_{m=1}^M \sum_{\boldsymbol{y} \in \mathcal{Y}^N} P(\boldsymbol{y}|m)\mathbb{1}\left\{\frac{e^{NT} \sum_{m' \neq m} P(\boldsymbol{y}|m')}{P(\boldsymbol{y}|m)} \geq 1\right\}$$

$$\leq \frac{1}{M} \sum_{m=1}^M \sum_{\boldsymbol{y} \in \mathcal{Y}^N} P(\boldsymbol{y}|m)\left(\frac{e^{NT} \sum_{m' \neq m} P(\boldsymbol{y}|m')}{P(\boldsymbol{y}|m)}\right)^\lambda \qquad (8)$$

$$\leq \frac{e^{N\lambda T}}{M} \sum_{m=1}^M \sum_{\boldsymbol{y} \in \mathcal{Y}^N} P^{\bar{\lambda}}(\boldsymbol{y}|m)\left(\sum_{m' \neq m} P(\boldsymbol{y}|m')\right)^\lambda$$

$$= e^{N\lambda T} \sum_{\boldsymbol{y} \in \mathcal{Y}^N} P^{\bar{\lambda}}(\boldsymbol{y}|m)\left(\sum_{m' \neq m} P(\boldsymbol{y}|m')\right)^\lambda \qquad (9)$$

where $0 \leq \lambda \leq 1$ is an arbitrary parameter, and $\bar{\lambda} \triangleq 1 - \lambda$. Similarly, for $\mathcal{E}_2$ we get that

$$\Pr\{\mathcal{E}_2\} \leq e^{-N\bar{\lambda}T} \sum_{\boldsymbol{y} \in \mathcal{Y}^N} P^{\bar{\lambda}}(\boldsymbol{y}|m)\left(\sum_{m' \neq m} P(\boldsymbol{y}|m')\right)^\lambda . \qquad (10)$$

Notice that the difference between these two expressions is given by a constant factor e^{-NT}, therefore, we will concentrate on achieving an upper bound on $\Pr\{\mathcal{E}_1\}$. It is important to note that starting this point on, our derivation is exponentially tight. Now, taking the expectation with respect to the code ensemble and using the fact that the encoding rule $\boldsymbol{U}(\boldsymbol{s}, m)$

(i.e., the process of choosing $\boldsymbol{u}$ given m and $\boldsymbol{s}$) is independent of all other encoding rules $\boldsymbol{U}(\boldsymbol{s}, m')$, we get that

$$\overline{\Pr\{\mathcal{E}_1\}} \le e^{N\lambda T} \sum_{\boldsymbol{y} \in \mathcal{Y}^N} \mathbb{E}_{\mathcal{C}} P^{\bar{\lambda}}(\boldsymbol{y}|m) \cdot \mathbb{E}_{\mathcal{C}} \left(\sum_{m' \ne m} P(\boldsymbol{y}|m') \right)^{\lambda},$$
(11)

where $\mathbb{E}_{\mathcal{C}}\{\cdot\}$ designates the expectation operator with respect to the code ensemble.

III. MAIN RESULTS

For a given joint distribution Q_{SUY} on $\mathcal{S} \times \mathcal{U} \times \mathcal{Y}$, a conditional distribution $W_{Y|SU} : \mathcal{S} \times \mathcal{U} \to \mathcal{Y}$ and a non-negative constant γ define

$$\mathcal{J}(Q_{SUY}, W_{Y|SU}, \gamma) \triangleq I_{Q_{SUY}}(U, S; Y) - \gamma \mathbb{E}_Q \ln W_{Y|SU},$$

where $\mathbb{E}_Q$ is the expectation operator associated with Q. For a given constant $K \ge 0$, a conditional distribution $P_{U|S}$ on $\mathcal{S} \times \mathcal{U}$ and a distribution P_Y on $\mathcal{Y}$, define the following set of distributions

$$\mathcal{G}(P_Y, K) \triangleq \Big\{ Q_{SU|Y} \mid$$
$$Q_{U|S} = P_{U|S},\ I_Q(U; S) - I_{Q \times P_Y}(U; Y) + K \ge 0 \Big\},$$
(12)

and define the complement of $\mathcal{G}(P_Y, K)$, denoted by $\mathcal{G}^c(P_Y, K)$, as

$$\mathcal{G}^c(P_Y, K) \triangleq \Big\{ Q_{SU|Y} \mid$$
$$Q_{U|S} = P_{U|S},\ I_Q(U; S) - I_{Q \times P_Y}(U; Y) + K < 0 \Big\}.$$
(13)

where it will be made clear from the text that $\mathcal{G}(P_Y, K)$ (and $\mathcal{G}^c(P_Y, K)$) is also a function of $P(U|S)$. The following theorem gives the main result of this work. The proof of the Theorem can be found in [14], [15].

Theorem 1:

$$\overline{\Pr\{\mathcal{E}_1\}} \le \exp \Bigg\{ - N \max_{0 \le \lambda \le 1} \Big[-\lambda T +$$
$$\max_{\substack{x:\mathcal{S} \times \mathcal{U} \to \mathcal{X} \\ P^*(U|S)}} \min_{P_Y \in \mathbb{P}^N(\mathcal{Y})} \Big(E_{ab}(\lambda, P_S, P_Y, \tilde{W}_{Y|SU}) +$$
$$E_{cd}(\lambda, R, P_S, P_Y, \tilde{W}_{Y|SU}) - H(P_Y) \Big) \Big] \Bigg\}.$$
(14)

where

$$E_{ab}(\lambda, P_S, P_Y, \tilde{W}_{Y|SU}) = \min \Bigg[\min_{Q_{SU|Y} \in \mathcal{G}(P_Y, 0)} \bar{\lambda} \Big(\mathcal{D}(Q_S \| P_S)$$
$$+ \mathcal{J}(Q_{SU|Y} \times P_Y, \tilde{W}_{Y|SU}, 1) \Big),$$
$$\min_{Q_{SU|Y} \in \mathcal{G}(P_Y, 0)^c} \Big(\bar{\lambda} \mathcal{D}(Q_S \| P_S) - \lambda I_{Q \times P_Y}(S; U, Y) +$$
$$\mathcal{J}(Q_{SU|Y} \times P_Y, \tilde{W}_{Y|SU}, \bar{\lambda}) \Big) \Bigg],$$
(15)

and

$$E_{cd}(\lambda, R, P_S, P_Y, \tilde{W}) = \min \Bigg[\min_{Q_{SU|Y} \in \mathcal{G}(P_Y, R)} \lambda \Big(\mathcal{D}(Q_S \| P_S)$$
$$- R + \mathcal{J}(Q_{SU|Y} \times P_Y, \tilde{W}_{Y|SU}, 1) \Big),$$
$$\min_{Q_{SU|Y} \in \mathcal{G}(P_Y, R)^c} \Big(\lambda \mathcal{D}(Q_S \| P_S) - \bar{\lambda} I_{Q \times P_Y}(S; U, Y) - R +$$
$$\mathcal{J}(Q_{SU|Y} \times P_Y, \tilde{W}_{Y|SU}, \lambda) \Big) \Bigg].$$
(16)

IV. ORDINARY DECODING RULE

The technique used in the proof of Theorem 1 can also be used to derive an achievable error exponent for ordinary decoding, i.e., a decoder which gives a single estimate for the transmitted message without the ability to declare "erase", when the same channel model and code-construction presented in Section II-A are used.

Again, by re-examining the derivation of Gallager's upper bound on the error probability presented in [16, Sec.2.4], it can be seen that the same results apply to channels with side-information available at the encoder non-causally where the only difference is that the probability of a received channel output $\boldsymbol{y}$ given a channel input $\boldsymbol{x}_m$ (denoted by $P(\boldsymbol{y}|\boldsymbol{x}_m)$) should be replaced by the probability of a received channel output $\boldsymbol{y}$ given a message index m (denoted by $P(\boldsymbol{y}|m)$). Hence, we get that

$$\Pr\{\mathcal{E}\} \le \frac{1}{M} \sum_{m=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}^N} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|m) \left[\sum_{m' \ne m} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|m') \right]^{\rho}$$
(17)

for $\rho > 0$. Taking the expectation with respect to the code ensemble, and using the fact that $\boldsymbol{U}(\boldsymbol{s}, m)$ is independent of all other encoding rules $\boldsymbol{U}(\boldsymbol{s}, m')$, we get

$$\overline{\Pr\{\mathcal{E}\}} \le \sum_{\boldsymbol{y} \in \mathcal{Y}^N} \mathbb{E}_{\mathcal{C}} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|m) \mathbb{E}_{\mathcal{C}} \left[\sum_{m' \ne m} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|m') \right]^{\rho}.$$
(18)

The following theorem provides an upper bound for the average decoding error.

Theorem 2:

$$\overline{\Pr\{\mathcal{E}\}} \le \exp \Bigg\{ - N \max_{\rho > 0} \Big[$$
$$\max_{\substack{x:\mathcal{S} \times \mathcal{U} \to \mathcal{X} \\ P^*(U|S)}} \min_{P_Y \in \mathbb{P}^N(\mathcal{Y})} \Big(E_{ab}(\frac{1}{1+\rho}, P_S, P_Y, \tilde{W}_{Y|SU}) +$$
$$\tilde{E}_{cd}(\lambda, R, P_S, P_Y, \tilde{W}_{Y|SU}) - H(P_Y) \Big) \Big] \Bigg\}.$$
(19)

where $E_{ab}\big(\frac{1}{1+\rho}, P_S, P_Y, \tilde{W}_{Y|SU}\big)$ is defined in eq.(15) and

$$
\begin{aligned}
\tilde{E}_{cd}(\rho, R, P_S, P_Y, \tilde{W}) = \min\Bigg[&\min_{Q_{SU|Y} \in \mathcal{G}(P_Y, R)} \rho\Big(\mathcal{D}(Q_S \| P_S) \\
&- R + \mathcal{J}\big(Q_{SU|Y} \times P_Y, \tilde{W}_{Y|SU}, \frac{1}{1+\rho}\big)\Big), \\
\min_{Q_{SU|Y} \in \mathcal{G}(P_Y, R)^c} &\Big(\rho\mathcal{D}(Q_S \| P_S) - \bar{\rho} I_{Q \times P_Y}(S; U, Y) - R + \\
&\mathcal{J}\big(Q_{SU|Y} \times P_Y, \tilde{W}_{Y|SU}, \frac{1}{1+\rho}\big)\Big)\Bigg].
\end{aligned}
\tag{20}
$$

Note that these two expressions $\mathbb{E}_{\mathcal{C}} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|m)$ and $\mathbb{E}_{\mathcal{C}} \big[\sum_{m' \neq m} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|m')\big]^{\rho}$ are already analyzed in the previous section.

V. Discussion and Future Work

Theorem 1 gives achievable error exponents of decoding with erasure/list decoding using the optimal decoding rule, originally proposed by Forney for DMCs, for channels with non-causal side information present at the transmitter. The main approach used during the analysis is the distance enumerator approach which was applied in [11] to obtain new error exponents of erasure/list decoding for DMCs. Theorem 2 is achieved using the same machinery used to devised Theorem 1. Both results originated from an optimal decoding rule (in the first results, the use of Forney's approach while in the second result Gallager's approach is used).

One more step needs to be done in order to generalize the above results. A generalization needs to be made to the way the channel input sequence $\boldsymbol{x}$ is chosen. While in [4], $\boldsymbol{x}$ was chosen randomly form the conditional type class $T^*_{\boldsymbol{x}|\boldsymbol{su}}$ induced by a conditional distribution $P^*(X|U, S)$ which can be optimized, in this work the encoder computes the channel input sequence $\boldsymbol{x}$ using a deterministic symbol-by-symbol function given the channel state sequence $\boldsymbol{s}$ and the auxiliary sequence $\boldsymbol{u}$ (see Section II-A). While in [13] such a deterministic mapping was sufficient to achieve the capacity of the channel it is not clear whether a deterministic mapping is optimal also in the error exponent regime.

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. Hof, I. Sason, and S. Shamai, "Performance bounds for erasure, list and feedback schemes with linear block codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3754–3778, Aug. 2010.

[4] E. Sabbag and N. Merhav, "Achievable error exponents for channels with side information – erasure and list decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5424–5431, Nov. 2010.

[5] 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.

[6] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 1-2, pp. 1–212, 2009.

[7] ——, "Relations between random coding exponents and the statistical physics of random codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 83–92, Jan. 2009.

[8] Y. Kaspi and N. Merhav, "Error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, submitted for publication, May 2009.

[9] R. Etkin, N. Merhav, and E. Ordentlich, "Error exponents of optimum decoding for the interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 40–56, Jan. 2010.

[10] A. S. Baruch and N. Merhav, "Exact random coding exponents for erasure decoding," in *Proc. Int. Symp. on Information Theory (ISIT'10)*, Austin TX, USA, Jun. 2010, pp. 260–264.

[11] N. Merhav, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4439–4447, Oct. 2008.

[12] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[13] 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.

[14] E. Sabbag and N. Merhav, "Achievable error exponents for channels with side information - erasure and list decoding," *In preparation. To be submitted to IEEE Trans. on Information Theory*, 2012.

[15] E. Sabbag, "Topics in channel reliability and error exponent analysis," Ph.D. dissertation, Technion - Israel Institute of Technology, 2011.

[16] A. J. Viterbi and J. K. Omura, *Principles of digital communication and coding*. New York: McGraw-Hill, 1979.