Ensemble-Tight Error Exponents for Mismatched Decoders

Jonathan Scarlett
University of Cambridge
jms265@cam.ac.uk

Alfonso Martinez
Universitat Pompeu Fabra
alfonso.martinez@ieee.org

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—This paper studies channel coding for discrete memoryless channels with a given (possibly suboptimal) decoding rule. Using upper and lower bounds on the random-coding error probability, the exponential behavior of three random-coding ensembles is characterized. The ensemble tightness of existing achievable error exponents is proven for the i.i.d. and constant-composition ensembles, and a new ensemble-tight error exponent is given for the cost-constrained i.i.d. ensemble. Connections are drawn between the ensembles under both mismatched decoding and maximum-likelihood decoding.**

I. Introduction

It is well known that random coding techniques can be used to prove the achievability part of Shannon's channel coding theorem, as well as characterizing the exponential behaviour of the best code for a range of rates under maximum-likelihood (ML) decoding [1]. In practice, however, ML decoding is often ruled out due to channel uncertainty and implementation constraints. In this paper, we consider the problem of *mismatched decoding* [2]–[8], in which the decoding rule is fixed and not necessarily optimal. In this setting, the following random-coding ensembles have been considered (see Section III for details):

1) the *i.i.d. ensemble*, in which each symbol of each codeword is generated independently;
2) the *constant-composition ensemble*, in which each codeword has the same empirical distribution;
3) the *cost-constrained i.i.d. ensemble*, in which each codeword satisfies a given cost constraint.

The i.i.d. ensemble can be used to prove the achievability of *Generalized Mutual Information* (GMI) [2], while the latter two ensembles can be used to prove the achievability of the higher *LM rate* [3], [4]. It is known that the LM rate is equal to the mismatched capacity in the case that the input alphabet is binary [5], but this is not true in general when the input is non-binary. It is therefore of interest to determine whether the weakness is due to the random-coding ensemble itself, or the bounding techniques used in the analysis.

This question was addressed in [6]–[8], where it was shown that the GMI and LM rate are tight with respect to the ensemble average for the i.i.d. and constant-composition ensembles respectively; see Section I-B for details. In this paper, we strengthen these results by obtaining ensemble-tight error exponents for each of the above ensembles. The analysis is performed in a unified fashion, and connections are drawn between the three ensembles.

This work has been supported in part by the European Research Council under ERC grant agreement 259663. A. Martinez received funding from the Ministry of Science and Innovation (Spain) under grant RYC-2011-08150 and from the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633.

A. System Setup

The input and output alphabets are denoted by $\mathcal{X}$ and $\mathcal{Y}$ respectively, and the channel $W(y|x)$ is assumed to be a discrete memoryless channel (DMC). We consider block coding, in which the codebook $\mathcal{C} = \{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(M)}\}$ is known at both the encoder and decoder. The encoder chooses a message m equiprobably from the set $\{1, ..., M\}$ and transmits the corresponding codeword $\boldsymbol{x}^{(m)}$. The decoder receives the vector $\boldsymbol{y}$ at the output of the channel, and forms the estimate

$$\hat{m} = \operatorname*{arg\,max}_{j\in\{1,...,M\}} \prod_{i=1}^{n} q(x_i^{(j)}, y_i) \tag{1}$$

where n is the length of each codeword and $x_i^{(j)}$ is the i-th entry of $\boldsymbol{x}^{(j)}$ (similarly for y_i). The function $q(x, y)$ is called the *decoding metric*, and is assumed to be non-negative. In the case of a tie, a random codeword achieving the maximum in (1) is selected. Throughout the paper, we write $q(\boldsymbol{x}, \boldsymbol{y})$ as a shorthand for $\prod_{i=1}^{n} q(x_i, y_i)$, and similarly for $W(\boldsymbol{y}|\boldsymbol{x})$. The *mismatched capacity* is defined to be the supremum of all rates $R = \frac{1}{n}\log M$ such that the error probability $p_e(\mathcal{C})$ can be made arbitrarily small for sufficiently large n.

An error exponent $E(R)$ is said to be achievable if there exists a sequence of codebooks of length n and rate R such that

$$\lim_{n\to\infty} -\frac{1}{n}\log p_e(\mathcal{C}) \geq E(R). \tag{2}$$

We let $\overline{p}_e$ denote the average error probability with respect to a given random-coding ensemble. The random-coding error exponent $E_r(R)$ is said to exhibit *ensemble tightness* if

$$\lim_{n\to\infty} -\frac{1}{n}\log \overline{p}_e = E_r(R). \tag{3}$$

B. Contributions and Previous Work

The GMI and LM rate are respectively defined as

$$I^{\mathrm{GMI}}(Q) \triangleq \sup_{s\geq 0} \mathbb{E}\left[\log \frac{q(X,Y)^s}{\mathbb{E}[q(\overline{X},Y)^s \,|\, Y]}\right] \tag{4}$$

$$I^{\mathrm{LM}}(Q) \triangleq \sup_{s\geq 0,a} \mathbb{E}\left[\log \frac{q(X,Y)^s e^{a(X)}}{\mathbb{E}[q(\overline{X},Y)^s e^{a(\overline{X})}\,|\,Y]}\right] \quad (5)$$

where $(X,Y,\overline{X}) \sim Q(x)W(y|x)Q(\overline{x})$, and $Q(x)$ is an arbitrary *input distribution*. For the i.i.d. ensemble with input distribution Q, it is known that $\overline{p}_e \to 0$ as $n \to \infty$ when $R < I^{\mathrm{GMI}}(Q)$, whereas $\overline{p}_e \to 1$ as $n \to \infty$ when $R > I^{\mathrm{GMI}}(Q)$ [2], [7]. Similarly, for the constant-composition ensemble with input distribution Q, it has been shown that $\overline{p}_e \to 0$ as $n \to \infty$ when $R < I^{\mathrm{LM}}(Q)$, whereas $\overline{p}_e \to 1$ as $n \to \infty$ when $R > I^{\mathrm{LM}}(Q)$ [6], [8].

While achievable error exponents exist in the literature for each of the aforementioned ensembles [2], [9], [10], we are not aware of any complete results on ensemble tightness. The ensemble tightness of error exponents in the matched regime was addressed in [11], but an extension of these techniques to the mismatched setting only proves ensemble tightness at low rates.

In this paper, we give upper and lower bounds to the random-coding error probability, and derive ensemble-tight error exponents for each ensemble. For the i.i.d. ensemble and constant-composition ensemble, our results prove the ensemble tightness of the achievable exponents presented in [2] and [9] respectively. The exponent for the cost-constrained i.i.d. ensemble appears to be new, and can be weakened to that of [10]. We draw connections between the three ensembles under both mismatched decoding and ML decoding.

C. Notation

The set of all probability distributions on an alphabet $\mathcal{A}$ is denoted by $\mathcal{P}(\mathcal{A})$, and the set of all empirical distributions on a vector in $\mathcal{A}^n$ (i.e. types) is denoted by $\mathcal{P}_n(\mathcal{A})$. The type of a vector $\boldsymbol{x}$ is denoted by $p_{\boldsymbol{x}}(\cdot)$. For a given $Q \in \mathcal{P}_n(\mathcal{A})$, the type class $T(Q)$ is defined to be the set of all sequences in $\mathcal{A}^n$ with type Q. We refer the reader to [12], [13] for an introduction to the method of types.

The probability of an event is denoted by $\mathbb{P}[\cdot]$, and the symbol $\sim$ means "distributed as". The marginals of a joint distribution $P_{XY}(x,y)$ are denoted by $P_X(x)$ and $P_Y(y)$. We write $P_X = \widetilde{P}_X$ to denote element-wise equality between two probability distributions on the same alphabet. Expectation with respect to a joint distribution $P_{XY}(x,y)$ is denoted by $\mathbb{E}_P[\cdot]$. When the associated probability distribution is understood from the context, the expectation is written as $\mathbb{E}[\cdot]$. Similarly, mutual information with respect to P_{XY} is written as $I_P(X;Y)$, or simply $I(X;Y)$ when the distribution is understood from the context. Given a distribution $Q(x)$ and a conditional distribution $W(y|x)$, we write $Q\times W$ to denote the joint distribution defined by $Q(x)W(y|x)$.

For two sequences $f(n)$ and $g(n)$, we write $f(n) \doteq g(n)$ if $\lim_{n\to\infty} \frac{1}{n}\log\frac{f(n)}{g(n)} = 0$, and similarly for $\dot{\leq}$ and $\dot{\geq}$. All logarithms have base e, and all rates are in units of nats except in the examples, where bits are used. We define $[c]^+ = \max\{0,c\}$, and denote the indicator function by $\mathbf{1}\{\cdot\}$.

II. Random-Coding Error Probability

In this section, we present a non-asymptotic analysis of the random-coding error probability $\overline{p}_e$ for an arbitrary codeword distribution $Q_{\boldsymbol{X}}(\boldsymbol{x})$. While it is possible to write an exact expression for $\overline{p}_e$ [14], its computation is usually infeasible even for moderate values of n. Similarly to [14], we can upper bound the error probability by assuming that ties are broken at random, and apply the union bound to obtain [15]

$$\begin{aligned}\overline{p}_e &\leq \mathrm{RCU}(n,M)\\ &\triangleq \mathbb{E}\big[\min\big\{1, (M-1)\mathbb{P}[q(\overline{\boldsymbol{X}},\boldsymbol{Y}) \geq q(\boldsymbol{X},\boldsymbol{Y})\,|\,\boldsymbol{X},\boldsymbol{Y}]\big\}\big]\end{aligned} \quad (6)$$

where $(\boldsymbol{X},\boldsymbol{Y},\overline{\boldsymbol{X}}) \sim Q_{\boldsymbol{X}}(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x})Q_{\boldsymbol{X}}(\overline{\boldsymbol{x}})$. This is the *random-coding union* (RCU) bound for mismatched decoders.

In order to lower bound the ensemble average error probability, we make use of a lower bound on the probability of a union of events due to de Caen, which states that [16]

$$\mathbb{P}\left[\bigcup_{i=1}^k A_i\right] \geq \sum_{i=1}^k \frac{\mathbb{P}[A_i]^2}{\sum_{j=1}^k \mathbb{P}[A_i \cap A_j]} \quad (7)$$

for an arbitrary sequence of probabilistic events $A_1,...,A_k$. In the case that the events are pairwise independent and identically distributed, we therefore obtain

$$\mathbb{P}\left[\bigcup_{i=1}^k A_i\right] \geq \frac{k\mathbb{P}[A_1]}{1+(k-1)\mathbb{P}[A_1]}. \quad (8)$$

Theorem 1. *The random-coding error probability for the mismatched decoder which resolves ties randomly satisfies*

$$\overline{p}_e \geq \mathrm{RCU_L}(n,M) \geq \frac{1}{4}\mathrm{RCU}(n,M) \quad (9)$$

where

$$\begin{aligned}&\mathrm{RCU_L}(n,M) \triangleq\\ &\frac{1}{2}\mathbb{E}\left[\frac{(M-1)\mathbb{P}[q(\overline{\boldsymbol{X}},\boldsymbol{Y}) \geq q(\boldsymbol{X},\boldsymbol{Y})\,|\,\boldsymbol{X},\boldsymbol{Y}]}{1+(M-2)\mathbb{P}[q(\overline{\boldsymbol{X}},\boldsymbol{Y}) \geq q(\boldsymbol{X},\boldsymbol{Y})\,|\,\boldsymbol{X},\boldsymbol{Y}]}\right]\\ &\quad+\frac{1}{2}\mathbb{E}\left[\frac{(M-1)\mathbb{P}[q(\overline{\boldsymbol{X}},\boldsymbol{Y}) > q(\boldsymbol{X},\boldsymbol{Y})\,|\,\boldsymbol{X},\boldsymbol{Y}]}{1+(M-2)\mathbb{P}[q(\overline{\boldsymbol{X}},\boldsymbol{Y}) > q(\boldsymbol{X},\boldsymbol{Y})\,|\,\boldsymbol{X},\boldsymbol{Y}]}\right]\end{aligned} \quad (10)$$

and $(\boldsymbol{X},\boldsymbol{Y},\overline{\boldsymbol{X}}) \sim Q_{\boldsymbol{X}}(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x})Q_{\boldsymbol{X}}(\overline{\boldsymbol{x}})$.

Proof: Consider a fixed codebook $\mathcal{C}$ with error probability $p_e(\mathcal{C})$. Let B_0 be the event that one or more codewords yield a strictly higher metric than the transmitted one, and let B_ℓ ($\ell \geq 1$) be the event that the transmitted codeword yields a metric which is equal highest with ℓ other codewords. We have

$$p_e(\mathcal{C}) = \mathbb{P}[B_0] + \sum_{\ell=1}^{M-1} \mathbb{P}[B_\ell] \frac{\ell}{\ell+1} \tag{11}$$

$$\geq \mathbb{P}[B_0] + \frac{1}{2} \sum_{\ell=1}^{M-1} \mathbb{P}[B_\ell] \tag{12}$$

$$= \frac{1}{2} p'_e(\mathcal{C}) + \frac{1}{2} \mathbb{P}[B_0] \tag{13}$$

where $p'_e(\mathcal{C}) \stackrel{\triangle}{=} \mathbb{P}[B_0] + \sum_{\ell=1}^{M-1} \mathbb{P}[B_\ell]$ is the error probability of a decoder which decodes ties as errors. Averaging (13) over the random-coding distribution, we obtain

$$\overline{p}_e \geq \frac{1}{2} \mathbb{P}\left[\bigcup_{i=2}^{M} \left\{ q(\boldsymbol{X}^{(i)}, \boldsymbol{Y}) \geq q(\boldsymbol{X}^{(1)}, \boldsymbol{Y}) \right\} \right] + \frac{1}{2} \mathbb{P}\left[\bigcup_{i=2}^{M} \left\{ q(\boldsymbol{X}^{(i)}, \boldsymbol{Y}) > q(\boldsymbol{X}^{(1)}, \boldsymbol{Y}) \right\} \right] \tag{14}$$

where $(\boldsymbol{X}^{(1)}, \boldsymbol{Y}) \sim Q_{\boldsymbol{X}}(\boldsymbol{x}) W(\boldsymbol{y}|\boldsymbol{x})$ and the $\boldsymbol{X}^{(i)}$ ($i \geq 2$) are generated according to $Q_{\boldsymbol{X}}(\boldsymbol{x})$, independently of $\boldsymbol{X}^{(1)}$ and $\boldsymbol{Y}$. The first inequality in (9) follows by writing each probability in (14) as an expectation given $\boldsymbol{X}^{(1)}$ and $\boldsymbol{Y}$, and applying the lower bound in (8). The second inequality in (9) follows by lower bounding (10) by the first of the two terms, replacing $M-2$ in the denominator with $M-1$, and applying the inequality $\frac{\alpha}{1+\alpha} \geq \frac{1}{2} \min\{1, \alpha\}$. ■

The complexity of the computation of $\mathrm{RCU_L}(n, M)$ is essentially identical to that of $\mathrm{RCU}(n, M)$. The second inequality in (9) shows that $\mathrm{RCU}(n, M)$ is ensemble-tight to within a factor of four, which will be useful for obtaining ensemble-tight error exponents in Section III.

We compare the upper and lower bounds numerically by considering the channel defined by the entries of the matrix

$$\begin{bmatrix} 1-2\delta_0 & \delta_0 & \delta_0 \\ \delta_1 & 1-2\delta_1 & \delta_1 \\ \delta_2 & \delta_2 & 1-2\delta_2 \end{bmatrix} \tag{15}$$

with $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$. The mismatched decoder chooses the codeword which is closest to $\boldsymbol{y}$ in terms of Hamming distance. For example, the decoding metric can be taken to be the entries of the matrix

$$\begin{bmatrix} 1-2\delta & \delta & \delta \\ \delta & 1-2\delta & \delta \\ \delta & \delta & 1-2\delta \end{bmatrix} \tag{16}$$

for any $\delta \in (0, \frac{1}{3})$. That is, the decoder uses a metric which is matched to a symmetric channel, but the true channel is asymmetric.

We set $\delta_0 = 0.01$, $\delta_1 = 0.05$ and $\delta_2 = 0.25$ and consider the ensemble in which each symbol of each codeword is generated independently according to $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Under these parameters, we have that $I^{\mathrm{GMI}}(Q) = 0.643$, $I^{\mathrm{LM}}(Q) = 0.728$ and $I(X; Y) = 0.763$ bits/use. Figure 1 plots $\mathrm{RCU}(n, M)$ and $\mathrm{RCU_L}(n, M)$ for $n = 50$, under both mismatched decoding and ML decoding. We observe

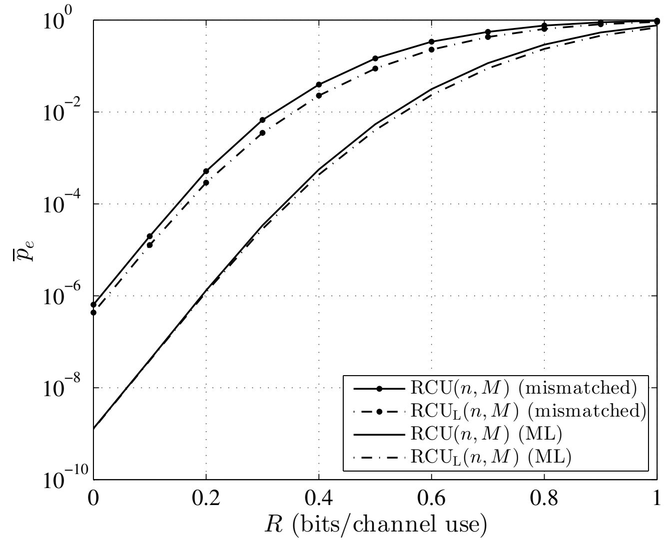

Figure 1. Random coding upper and lower bounds for the channel defined in (15) with $n = 50$, $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The mismatched decoder uses the minimum Hamming distance metric given in (16).

a very close match between the upper and lower bounds across all rates, particularly in the case of ML decoding. The slightly larger gap in the mismatched case is due to an increased probability of decoding ties, arising from the fact that a simpler decoding metric is being used.

III. Random-Coding Error Exponents

In this section, we consider three families of the random-coding distribution $Q_{\boldsymbol{X}}(\boldsymbol{x})$, each of which depends on an input distribution $Q(x)$.

1) The *i.i.d. ensemble* is given by

$$Q_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} Q(x_i). \tag{17}$$

In words, each symbol of each codeword is generated independently according to Q.

2) The *constant-composition ensemble* is given by

$$Q_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{|T(Q_n)|} \mathbb{1}\{\boldsymbol{x} \in T(Q_n)\} \tag{18}$$

where Q_n is the most probable type under Q. That is, each codeword is generated uniformly over the type class $T(Q_n)$, and hence each codeword has the same composition.

3) The *cost-constrained i.i.d. ensemble* is given by

$$Q_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\mu_n} \prod_{i=1}^{n} Q(x_i) \mathbb{1}\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} a(x_i) - \phi_a \right| \leq \frac{\delta}{n} \right\} \tag{19}$$

where $a(x)$ is a *cost function*, $\phi_a \stackrel{\triangle}{=} \mathbb{E}_Q[a(X)]$, δ is a positive constant which does not vary with n, and μ_n is a normalizing constant. Roughly speaking, each codeword is generated according to an i.i.d. distribution *conditioned* on the empirical mean of $a(x)$ being very close to the true mean. The cost function a should *not* be viewed as being chosen to meet a system constraint (e.g. power limitations). Rather, it is introduced in order to improve the performance of the random-coding ensemble; see [7], [10] for details.

We can rewrite each of the random-coding distributions in (17)–(19) as an i.i.d. distribution conditioned on the empirical distribution of $\boldsymbol{x}$ being in a particular set of types. To this end, we introduce the general ensemble defined by

$$Q_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\mu_n'} \prod_{i=1}^{n} Q(x_i) \mathbb{1}\{p_{\boldsymbol{x}} \in \mathcal{G}_n\} \tag{20}$$

where $\mathcal{G}_n \in \mathcal{P}_n(\mathcal{X})$ is the set of possible codeword types and μ_n' is a normalizing constant. The i.i.d. ensemble is recovered by setting $\mathcal{G}_n = \mathcal{P}_n(\mathcal{X})$, the constant-composition ensemble is recovered by setting $\mathcal{G}_n = \{Q_n\}$, and the cost-constrained i.i.d. ensemble is recovered by setting $\mathcal{G}_n = \{P_X : |\mathbb{E}_P[a(X)] - \phi_a| \le \frac{\delta}{n}\}$.

We determine the exponential behaviour of the general ensemble in (20) using known properties of types, analogous to the analysis of the ML decoder given in [13]. We define the sets

$$\mathcal{S}_n(\mathcal{G}_n) \triangleq \big\{P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) \,:\, P_X \in \mathcal{G}_n\big\} \tag{21}$$

$$\begin{aligned}\mathcal{T}_n(P_{XY}, \mathcal{G}_n) \triangleq \big\{&\widetilde{P}_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) \,:\, \widetilde{P}_X \in \mathcal{G}_n, \\ &\widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]\big\}.\end{aligned} \tag{22}$$

Roughly speaking, $\mathcal{S}_n$ is the set of possible types of $(\boldsymbol{X}^{(m)}, \boldsymbol{Y})$, and $\mathcal{T}_n$ is the set of types of $(\boldsymbol{X}^{(m')}, \boldsymbol{Y})$ which lead to errors given that $(\boldsymbol{X}^{(m)}, \boldsymbol{Y}) \in T(P_{XY})$, where m is the transmitted message and m' is a different message.

Theorem 2. *Suppose $\mathcal{G}_n$ is such that $\mathbb{P}[\boldsymbol{X}' \in \mathcal{G}_n] \doteq 1$, where $\boldsymbol{X}' \sim \prod_{i=1}^n Q(x_i)$. Then the random-coding error probability for the random-coding ensemble in* (20) *satisfies*

$$\overline{p}_e \doteq \exp\big(-nE_{r,n}(Q,R,\mathcal{G}_n)\big) \tag{23}$$

where

$$\begin{aligned}E_{r,n}(Q,R,\mathcal{G}_n) \triangleq &\min_{P_{XY} \in \mathcal{S}_n(\mathcal{G}_n)} \min_{\widetilde{P}_{XY} \in \mathcal{T}_n(P_{XY},\mathcal{G}_n)} \\ &D(P_{XY} \| Q \times W) + \Big[D(\widetilde{P}_{XY} \| Q \times \widetilde{P}_Y) - R\Big]^+.\end{aligned} \tag{24}$$

Proof: From (6) and (9), we have

$$\overline{p}_e \doteq \mathrm{RCU}(n,M) = \mathbb{E}\left[\psi(\boldsymbol{X},\boldsymbol{Y})\right] \tag{25}$$

where

$$\psi(\boldsymbol{x},\boldsymbol{y}) \triangleq \min\big\{1, (M-1)\mathbb{P}\big[q(\overline{\boldsymbol{X}},\boldsymbol{y}) \ge q(\boldsymbol{x},\boldsymbol{y})\big]\big\}. \tag{26}$$

Let P_{XY} denote the joint type of $(\boldsymbol{x},\boldsymbol{y})$. Since $\psi(\boldsymbol{x},\boldsymbol{y})$ depends only on P_{XY}, we write $\psi(P_{XY}) \triangleq \psi(\boldsymbol{x},\boldsymbol{y})$. From the definition of $\mathcal{T}_n(P_{XY},\mathcal{G}_n)$, we have that $q(\overline{\boldsymbol{X}},\boldsymbol{y}) \ge q(\boldsymbol{x},\boldsymbol{y})$ if and only if $(\overline{\boldsymbol{X}},\boldsymbol{y}) \in \mathcal{T}_n(P_{XY},\mathcal{G}_n)$, and hence

$$\begin{aligned}\psi(P_{XY}) = \min\Big\{1, \\ (M-1) \sum_{\widetilde{P}_{XY} \in \mathcal{T}_n(P_{XY},\mathcal{G}_n)} \mathbb{P}\big[(\overline{\boldsymbol{X}},\boldsymbol{y}) \in T(\widetilde{P}_{XY})\big]\Big\}.\end{aligned} \tag{27}$$

From (20), the distribution of $\overline{\boldsymbol{X}}$ is the same as that of $\boldsymbol{X}' \sim \prod_{i=1}^n Q(x_i)$ conditioned on the event that $\boldsymbol{X}' \in \mathcal{G}_n$. Hence, using the assumption that $\mathbb{P}[\boldsymbol{X}' \in \mathcal{G}_n] \doteq 1$, we obtain

$$\begin{aligned}\psi(P_{XY}) \doteq \min\big\{1, \\ (M-1) \sum_{\widetilde{P}_{XY} \in \mathcal{T}_n(P_{XY},\mathcal{G}_n)} \mathbb{P}\big[(\boldsymbol{X}',\boldsymbol{y}) \in T(\widetilde{P}_{XY})\big]\big\}\end{aligned} \tag{28}$$

$$\begin{aligned}\doteq \max_{\widetilde{P}_{XY} \in \mathcal{T}_n(P_{XY},\mathcal{G}_n)} \min\big\{1, \\ (M-1)\exp\big(-nD(\widetilde{P}_{XY} \| Q \times \widetilde{P}_Y)\big\}\end{aligned} \tag{29}$$

where (29) follows from the property of types in [13, Eq. (18)], and the fact that the number of joint types is polynomial in n.

Expanding the expectation in (25), we obtain

$$\overline{p}_e \doteq \sum_{P_{XY} \in \mathcal{S}_n(\mathcal{G}_n)} \mathbb{P}\left[(\boldsymbol{X},\boldsymbol{Y}) \in T(P_{XY})\right] \psi(P_{XY}) \tag{30}$$

and a nearly identical argument to (27)–(29) yields

$$\overline{p}_e \doteq \max_{P_{XY} \in \mathcal{S}_n(\mathcal{G}_n)} \exp\big(-nD(P_{XY} \| Q \times W)\big)\psi(P_{XY}). \tag{31}$$

The proof is concluded by substituting (29) into (31). ■

Using Theorem 2, we obtain ensemble-tight error exponents for the ensembles defined in (17)–(19). Specifically, defining the sets

$$\mathcal{S}^{\mathrm{iid}} \triangleq \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \tag{32}$$

$$\mathcal{S}^{\mathrm{cc}}(Q) \triangleq \big\{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \,:\, P_X = Q\big\} \tag{33}$$

$$\mathcal{S}^{\mathrm{cost}}(a) \triangleq \big\{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \,:\, \mathbb{E}_P[a(X)] = \phi_a\big\} \tag{34}$$

and

$$\begin{aligned}\mathcal{T}^{\mathrm{iid}}(P_{XY}) \triangleq \big\{&\widetilde{P}_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \,: \\ &\widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]\big\}\end{aligned} \tag{35}$$

$$\begin{aligned}\mathcal{T}^{\mathrm{cc}}(P_{XY}) \triangleq \big\{&\widetilde{P}_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \,:\, \widetilde{P}_X = P_X, \\ &\widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]\big\}\end{aligned} \tag{36}$$

$$\begin{aligned}\mathcal{T}^{\mathrm{cost}}(P_{XY}, a) \triangleq \big\{&\widetilde{P}_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \,:\, \mathbb{E}_{\widetilde{P}}[a(X)] = \phi_a, \\ &\widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]\big\}\end{aligned} \tag{37}$$

we obtain the following corollary.

Corollary 3. *The random-coding error exponents for the ensembles defined in* (17)–(19) *are respectively given by*

$$E_r^{\mathrm{iid}}(Q,R) \triangleq \min_{P_{XY}\in\mathcal{S}^{\mathrm{iid}}} \min_{\widetilde{P}_{XY}\in\mathcal{T}^{\mathrm{iid}}(P_{XY})} D(P_{XY}\|Q\times W) + \Big[D(\widetilde{P}_{XY}\|Q\times\widetilde{P}_Y) - R\Big]^+ \quad (38)$$

$$E_r^{\mathrm{cc}}(Q,R) \triangleq \min_{P_{XY}\in\mathcal{S}^{\mathrm{cc}}(Q)} \min_{\widetilde{P}_{XY}\in\mathcal{T}^{\mathrm{cc}}(P_{XY})} D(P_{XY}\|Q\times W) + \Big[I_{\widetilde{P}}(X;Y) - R\Big]^+ \quad (39)$$

$$E_r^{\mathrm{cost}}(Q,R,a) \triangleq \min_{P_{XY}\in\mathcal{S}^{\mathrm{cost}}(a)} \min_{\widetilde{P}_{XY}\in\mathcal{T}^{\mathrm{cost}}(P_{XY},a)} D(P_{XY}\|Q\times W) + \Big[D(\widetilde{P}_{XY}\|Q\times\widetilde{P}_Y) - R\Big]^+. \quad (40)$$

Proof: The assumption $\mathbb{P}[\boldsymbol{X}'\in\mathcal{G}_n]\doteq 1$ in Theorem 2 holds trivially for the i.i.d. ensemble, and was shown to hold in [12, Eq. (2.6)] and [10, Eq. (88)] for the constant-composition ensemble and cost-constrained i.i.d. ensemble respectively. Since any probability distribution can be approximated arbitrarily well by a type for sufficiently large n, the minimizations over types in (24) can be replaced by minimizations over all probability distributions [12]. Similarly, the constraint $P_X = Q_n$ for the constant-composition ensemble can be replaced by $P_X = Q$, and the constraint $|\mathbb{E}_P[a(X)]-\phi_a|\le\frac{\delta}{n}$ for the cost-constrained i.i.d. ensemble can be replaced by $\mathbb{E}_P[a(X)]=\phi_a$, regardless of the value of δ. Similar arguments apply to the constraints on $\widetilde{P}_X$ for the constant-composition and cost-constrained i.i.d. ensembles. ■

The optimization problems in (38)–(40) are all convex for a fixed input distribution and cost function. Using the method of Lagrange duality [17], each exponent can be written in an alternative form.

Theorem 4. *The error exponents in* (38)–(40) *can be expressed as*

$$E_r^{\mathrm{iid}}(Q,R) = \max_{\rho\in[0,1]} E_0^{\mathrm{iid}}(Q,\rho) - \rho R \quad (41)$$

$$E_r^{\mathrm{cc}}(Q,R) = \max_{\rho\in[0,1]} E_0^{\mathrm{cc}}(Q,\rho) - \rho R \quad (42)$$

$$E_r^{\mathrm{cost}}(Q,R,a) = \max_{\rho\in[0,1]} E_0^{\mathrm{cost}}(Q,\rho,a) - \rho R \quad (43)$$

where

$$E_0^{\mathrm{iid}}(Q,\rho) \triangleq \sup_{s\ge 0} -\log\mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X},Y)^s\,|\,Y\big]}{q(X,Y)^s}\right)^\rho\right] \quad (44)$$

$$E_0^{\mathrm{cc}}(Q,\rho) \triangleq \sup_{s\ge 0,a} \mathbb{E}\left[-\log\mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X},Y)^s e^{a(\overline{X})}\,|\,Y\big]}{q(X,Y)^s e^{a(X)}}\right)^\rho \,\middle|\, X\right]\right] \quad (45)$$

$$E_0^{\mathrm{cost}}(Q,\rho,a) \triangleq \sup_{s\ge 0,r,\overline{r}} -\log\mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X},Y)^s e^{\overline{r}(a(\overline{X})-\phi_a)}\,|\,Y\big]}{q(X,Y)^s e^{r(a(X)-\phi_a)}}\right)^\rho\right] \quad (46)$$

and $(X,Y,\overline{X})\sim Q(x)W(y|x)Q(\overline{x})$.

Proof: The proofs are similar for each of the three ensembles, so we provide a sketch only for E_r^{cc}. Applying $[a]^+ = \max_{\rho\in[0,1]}\rho a$ to (39) and using Fan's minimax theorem [18], we obtain

$$E_r^{\mathrm{cc}}(Q,R) = \max_{\rho\in[0,1]} \hat{E}_0^{\mathrm{cc}}(Q,\rho) - \rho R \quad (47)$$

where

$$\hat{E}_0^{\mathrm{cc}}(Q,\rho) \triangleq \min_{P_{XY}\in\mathcal{S}^{\mathrm{cc}}(Q)} \min_{\widetilde{P}_{XY}\in\mathcal{T}^{\mathrm{cc}}(P_{XY})} D(P_{XY}\|Q\times W) + \rho I_{\widetilde{P}}(X;Y). \quad (48)$$

It remains to show via Lagrange duality that $\hat{E}_0^{\mathrm{cc}}(Q,\rho) = E_0^{\mathrm{cc}}(Q,\rho)$. We first fix P_{XY} and consider the problem

$$\min_{\widetilde{P}_{XY}\in\mathcal{T}^{\mathrm{cc}}(P_{XY})} I_{\widetilde{P}}(X;Y) \quad (49)$$

the Lagrange dual of which is given by [6]

$$\sup_{s\ge 0,a} \sum_{x,y} P_{XY}(x,y)\log\frac{q(x,y)^s e^{a(x)}}{\sum_{\overline{x}} P_X(\overline{x})q(\overline{x},y)^s e^{a(\overline{x})}} \quad (50)$$

where s and $a(x)$ are Lagrange multipliers. Substituting (50) into (48) yields a min-sup problem, where the minimization is over $P_{XY}\in\mathcal{S}(P)$ and the supremum is over $s\ge 0$ and a. Using Fan's minimax theorem [18], the order of these optimizations can be swapped. The proof is concluded by forming the Lagrange dual problem of the resulting optimization over P_{XY}. ■

The expressions in (39) and (41) appear in [9] and [2] respectively, though both derivations are different to ours. To our knowledge, the alternative expressions in (38) and (42) have not appeared previously, and the exponent E_r^{cost} is new. We note that the function $a(x)$ represents different quantities for the constant-composition ensemble and cost-constrained i.i.d. ensemble. In the former case, a arises as a mathematical optimization parameter, whereas in the latter case, a is a design parameter for the random-coding ensemble.

The exponents in (41)–(43) can be derived directly, rather than via Lagrange duality. The derivation of (41) is presented in [2], and the expression in (42) follows by combining the achievable error exponent of [19] with the fact that under constant-composition codes, the metric $q(x,y)$ is equivalent to the metric $q(x,y)^s e^{a(x)}$ for any $s\ge 0$ and a [3].

The direct derivation of (43) is similar to that of [10], where it was shown that an achievable error exponent for the cost-constrained i.i.d. ensemble is given by

$$E_r^{\mathrm{cost}'}(Q,R,a) \triangleq \max_{\rho\in[0,1]} E_0^{\mathrm{cost}'}(Q,\rho,a) - \rho R \quad (51)$$

$$E_0^{\mathrm{cost}'}(Q,\rho,a) \triangleq \sup_{s\geq 0} -\log \mathbb{E}\left[\left(\frac{\mathbb{E}[q(\overline{X},Y)^s e^{a(\overline{X})}\,|\,Y]}{q(X,Y)^s e^{a(X)}}\right)^{\rho}\right]. \quad (52)$$

Roughly speaking, the additional factor of $\frac{e^{a(\overline{x})}}{e^{a(x)}}$ in (52) compared to the i.i.d. ensemble in (44) arises from the fact that $e^{\sum_i a(\overline{x}_i)}$ is close to $e^{\sum_i a(x_i)}$ for any codewords $\boldsymbol{x}$ and $\overline{\boldsymbol{x}}$. For the function E_0^{cost} in (46), the additional factor $e^{r(a(x)-\phi_a)}$ arises from the fact that $e^{\sum_i a(x_i)}$ is close to $e^{n\phi_a}$, and similarly for the factor $e^{\overline{r}(a(\overline{x})-\phi_a)}$. We will see that the refined exponent E_r^{cost} improves on $E_r^{\mathrm{cost}'}$ in general.

While none of the above direct derivations prove ensemble tightness, they each have the advantage of extending immediately to continuous alphabets after replacing the appropriate sums by integrals, except that the input alphabet must be finite for the constant-composition ensemble.

A. Connections Between the Error Exponents

The constraints on P_{XY} and $\widetilde{P}_{XY}$ in (41)–(43) are given by the sets defined in (32)–(37). Since the constraint $\mathbb{E}_P[a(X)] = \phi_a$ holds by definition when $P_X = Q$, we have that $\mathcal{S}^{\mathrm{iid}} \subseteq \mathcal{S}^{\mathrm{cost}}(a) \subseteq \mathcal{S}^{\mathrm{cc}}(Q)$ for any given cost function a and input distribution Q. A similar observation applies to the constraints on $\widetilde{P}_X$, and it follows that

$$E_r^{\mathrm{iid}}(Q,R) \leq E_r^{\mathrm{cost}}(Q,R,a) \leq E_r^{\mathrm{cc}}(Q,R). \quad (53)$$

This indicates that the constant-composition ensemble yields the best error exponent of the three ensembles under consideration. Furthermore, by setting $r = \overline{r} = 1$ in (46), we obtain the inequality

$$E_r^{\mathrm{cost}'}(Q,R,a) \leq E_r^{\mathrm{cost}}(Q,R,a) \quad (54)$$

where strict inequality is possible. In the case that $a(x)$ does not depend on x, we obtain

$$E_r^{\mathrm{cost}}(Q,R,a) = E_r^{\mathrm{cost}'}(Q,R,a) = E_r^{\mathrm{iid}}(Q,R) \quad (55)$$

and hence we have in general that

$$E_r^{\mathrm{iid}}(Q,R) \leq \sup_a E_r^{\mathrm{cost}'}(Q,R,a). \quad (56)$$

The following theorem gives the connection between $E_r^{\mathrm{cost}'}$ and E_r^{cc}, and shows that the two are equal after optimization over Q and a. This result is analogous to a connection between the i.i.d. ensemble and constant-composition ensemble under ML decoding [19].

Theorem 5. *$E_0^{\mathrm{cc}}(Q,\rho)$ can be expressed as*

$$E_0^{\mathrm{cc}}(Q,\rho) = \max_{\widetilde{Q}\in\mathcal{P}(\mathcal{X})} \sup_a E_0^{\mathrm{cost}'}(\widetilde{Q},\rho,a) - (1+\rho)D(Q\|\widetilde{Q}) \quad (57)$$

Consequently,

$$\max_{Q\in\mathcal{P}(\mathcal{X})} E_r^{\mathrm{cc}}(Q,R) = \max_{Q\in\mathcal{P}(\mathcal{X})} \sup_a E_r^{\mathrm{cost}'}(Q,R,a). \quad (58)$$

Proof: Since the supremum over a in (45) is over all real-valued functions on $\mathcal{X}$, an equivalent expression is obtained by defining $\widetilde{a}(x)$ such that $e^{\widetilde{a}(x)} = e^{a(x)}\frac{Q(x)}{\widetilde{Q}(x)}$ for some $\widetilde{Q}(x)$,[1] and instead taking the supremum over $\widetilde{a}$. Simple algebraic manipulations yield

$$E_0^{\mathrm{cc}}(Q,\rho) = \sup_{s\geq 0,\widetilde{a}} -\sum_x Q(x)\log\left(\frac{\widetilde{Q}(x)}{Q(x)}\sum_y W(y|x) \times \left(\frac{\sum_{\overline{x}}\widetilde{Q}(\overline{x})q(\overline{x},y)e^{\widetilde{a}(\overline{x})}}{q(x,y)^s e^{\widetilde{a}(x)}}\right)^{\rho}\right) - (1+\rho)D(Q\|\widetilde{Q}) \quad (59)$$

$$\geq \sup_{s\geq 0,\widetilde{a}} -\log\left(\sum_{x,y}\widetilde{Q}(x)W(y|x) \times \left(\frac{\sum_{\overline{x}}\widetilde{Q}(\overline{x})q(\overline{x},y)e^{\widetilde{a}(\overline{x})}}{q(x,y)^s e^{\widetilde{a}(x)}}\right)^{\rho}\right) - (1+\rho)D(Q\|\widetilde{Q}) \quad (60)$$

where (60) follows from Jensen's inequality.

It remains to show that equality holds in (60) after maximizing over $\widetilde{Q}$. First, by analyzing the Karush-Kuhn-Tucker (KKT) conditions for the optimization problem over $\widetilde{Q}$, s and $\widetilde{a}$ associated with (60), it can be shown that

$$\frac{\widetilde{Q}(x)}{Q(x)}\sum_y W(y|x)\left(\sum_{\overline{x}}\widetilde{Q}(\overline{x})\left(\frac{q(\overline{x},y)}{q(x,y)}\right)^s\frac{e^{\widetilde{a}(\overline{x})}}{e^{\widetilde{a}(x)}}\right)^{\rho} \quad (61)$$

is constant for all x such that $Q(x) > 0$ under the optimal parameters, implying equality in (60). The proof is concluded by showing that the resulting values of s and a (the latter being computed using $\widetilde{Q}$ and $\widetilde{a}$) also satisfy the KKT conditions for the optimization problem over s and a associated with (45). Details are omitted for the sake of readability. The equality in (58) follows immediately from (57) and the fact that $D(Q\|\widetilde{Q}) \geq 0$ with equality if $Q = \widetilde{Q}$. ∎

It follows from (53) and Theorem 5 that $E_r^{\mathrm{cost}'}$ is tight with respect to the ensemble average when the optimal values of both Q and a are used. Hence, although E_r^{cost} is a tighter error exponent than $E_r^{\mathrm{cost}'}$ in general, it does not improve on the *best* achievable error exponent using cost-constrained i.i.d. random coding. Furthermore, both exponents can be used to prove the achievability of the LM rate and no better. However, the refined exponent E_r^{cost} is useful in the case that one does not have complete freedom in choosing Q and a, or when exact optimization over each is not feasible. For example, if the codebook designer does not know the channel, then the objective in (52) cannot be computed in order to perform the optimization.

The following theorem gives two further connections between the error exponents in the case of ML decoding.

Theorem 6. *If $q(x,y) = W(y|x)$ then*

$$\sup_a E_r^{\mathrm{cost}'}(Q,R,a) = E_r^{\mathrm{iid}}(Q,R) \quad (62)$$

[1] We assume that $\widetilde{Q}(x) > 0$ wherever $Q(x) > 0$, since all other choices of $\widetilde{Q}$ make the objective in (57) equal to $-\infty$ and hence cannot achieve the maximum.

and

$$\sup_{a} E_r^{\text{cost}}(Q,R,a) = E_r^{\text{cc}}(Q,R) \tag{63}$$

for any given Q and R.

Proof: We obtain (62) by optimizing the objective in (44) over s, and optimizing the objective in (52) over s and a. It was shown in [1, Ex. 5.6] that the optimal value of s in (44) is equal to $\frac{1}{1+\rho}$. Following the same steps, we obtain that that the optimal value of s is in (52) is also equal to $\frac{1}{1+\rho}$, and the optimal cost function $a(x)$ does not depend on x. Combining these results, we obtain (62).

From (53), it suffices to prove that $\sup_a E_r^{\text{cost}} \ge E_r^{\text{cc}}$ in order to prove (63). To this end, we will show that $\sup_a E_0^{\text{cost}}(Q,\rho,a) \ge E_0^{\text{cc}}(Q,\rho)$ for all $\rho \in [0,1]$. The case $\rho = 0$ is trivial, so we assume that $\rho > 0$. We set $\overline{r} = 1$ and $r = \frac{-1}{\rho}$ and write the expectation inside the logarithm of (46) as

$$\sum_{x,y} Q(x)W(y|x)\left(\frac{\sum_{\overline{x}} Q(\overline{x})q(\overline{x},y)^s e^{a(\overline{x})}}{q(x,y)^s e^{\phi_a}}\right)^{\rho} \frac{e^{a(x)}}{e^{\phi_a}}. \tag{64}$$

We assume without loss of generality that $Q(x) > 0$ for all x. Introducing the distribution $\widetilde{Q}(x) = \frac{Q(x)e^{a(x)}}{\mathbb{E}_Q[e^{a(X)}]}$, we can write (64) as

$$\left(\frac{\mathbb{E}_Q[e^{a(X)}]}{e^{\phi_a}}\right)^{1+\rho} \sum_{x,y} \widetilde{Q}(x)W(y|x)\left(\frac{\sum_{\overline{x}} \widetilde{Q}(\overline{x})q(\overline{x},y)^s}{q(x,y)^s}\right)^{\rho}. \tag{65}$$

The summation in (65) coincides with the expectation inside the logarithm of (44), and with some simple algebraic manipulation it can be shown that

$$\frac{\mathbb{E}_Q[e^{a(X)}]}{e^{\phi_a}} = \exp\big(D(Q\|\widetilde{Q})\big). \tag{66}$$

Substituting (65) and (66) into (46) and noting that a suitable choice of a can yield any distribution $\widetilde{Q}$ such that $\widetilde{Q}(x) > 0$ for all x, we obtain

$$\sup_{a} E_0^{\text{cost}}(Q,\rho,a) \ge \max_{\widetilde{Q}} E_0^{\text{iid}}(\widetilde{Q},\rho) - (1+\rho)D(Q\|\widetilde{Q}). \tag{67}$$

Taking account of (62) and Theorem 5, the right-hand side of (67) is equal to $E_0^{\text{cc}}(Q,\rho)$, and the proof is complete. ■

Under ML decoding, E_r^{iid} is simply Gallager's random-coding error exponent [1], while E_r^{cc} is Csiszár's random-coding error exponent for constant-composition codes [12]. We have thus shown that using an optimized cost function a, we can achieve Csiszár's exponent in the matched setting without using constant-composition codes. However, the analysis thus far does not give a precise connection between E_r^{cost} and E_r^{cc} when the decoding metric differs from ML.

B. Multiple Cost Constraints

In this subsection, we outline how the exponent E_r^{cc} can be achieved in the mismatched setting using cost-constrained i.i.d. random coding. To this end, we introduce the cost-constrained i.i.d. ensemble with L cost constraints, given by

$$Q_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\mu_n}\prod_{i=1}^{n} Q(x_i)\mathbf{1}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} a_l(x_i) - \phi_l\right| \le \frac{\delta_l}{n},\, l = 1,...,L\right\} \tag{68}$$

where for each $l \in \{1,...,L\}$, a_l is a cost function, $\phi_l \triangleq \mathbb{E}_Q[a_l(X)]$ and δ_l is a positive constant. By using the multidimensional central limit theorem and extending the analysis of [10, Eq. (88)], it can be shown that the normalizing constant μ_n decays to zero sub-exponentially, i.e. $\mu_n \doteq 1$. Using this result and following the analysis of the case $L = 1$ at the beginning of this section, we obtain the exponent

$$E_r^{\text{cost}}(Q,R,\{a_l\}) = \max_{\rho\in[0,1]} E_0^{\text{cost}}(Q,\rho,\{a_l\}) - \rho R \tag{69}$$

where

$$E_0^{\text{cost}}(Q,\rho,\{a_l\}) \triangleq \sup_{s\ge 0,\{r_l\},\{\overline{r}_l\}} - \log \mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X},Y)^s e^{\sum_{l=1}^{L} \overline{r}_l(a_l(\overline{X})-\phi_l)} \,|\, Y\big]}{q(X,Y)^s e^{\sum_{l=1}^{L} r_l(a_l(X)-\phi_l)}}\right)^{\rho}\right]. \tag{70}$$

Here we write $\{a_l\}$ as a shorthand for $\{a_1,...,a_L\}$, and similarly for $\{r_l\}$ and $\{\overline{r}_l\}$.

The constant-composition ensemble is in fact a special case of the ensemble in (68), since it is obtained by setting $L = |\mathcal{X}| - 1$ and $\delta_l < 1$ for all l, and choosing the cost functions $a_1 = (1,0,...,0)$, $a_2 = (0,1,0,...,0)$, etc. We will show, however, that the exponent E_r^{cc} can be recovered using only two cost functions, regardless of the cardinality of $\mathcal{X}$. Setting $L = 2$ and choosing $r_1 = \overline{r}_1 = \overline{r}_2 = 1$ and $r = \frac{-1}{\rho}$, the expectation in (70) becomes

$$\sum_{x,y} Q(x)W(y|x)\left(\frac{\sum_{\overline{x}} Q(\overline{x})q(\overline{x},y)^s e^{a_1(\overline{x})+a_2(\overline{x})}}{q(x,y)^s e^{a_1(x)+\phi_2}}\right)^{\rho} \frac{e^{a_2(x)}}{e^{\phi_2}}. \tag{71}$$

Defining $\widetilde{Q}(x) = \frac{Q(x)e^{a_2(x)}}{\mathbb{E}_Q[e^{a_2(X)}]}$ and following identical steps to (65)–(67), we obtain

$$\sup_{a_1,a_2} E_0^{\text{cost}}(Q,\rho,\{a_1,a_2\}) \ge \max_{\widetilde{Q}} \sup_{a_1} E_0^{\text{cost}'}(\widetilde{Q},\rho,a_1) - (1+\rho)D(Q\|\widetilde{Q}). \tag{72}$$

By Theorem 5, the right-hand side of (72) is equal to $E_0^{\text{cc}}(Q,\rho)$, and we have thus recovered the exponent E_r^{cc}.

C. Numerical Results

In this subsection, we plot the exponents for the channel defined in (15) under both the minimum Hamming distance and ML decoding metrics, again using the parameters $\delta_0 = 0.01$, $\delta_1 = 0.05$ and $\delta_2 = 0.25$. We set $Q = (0.1, 0.3, 0.6)$, which we have intentionally chosen suboptimally to highlight

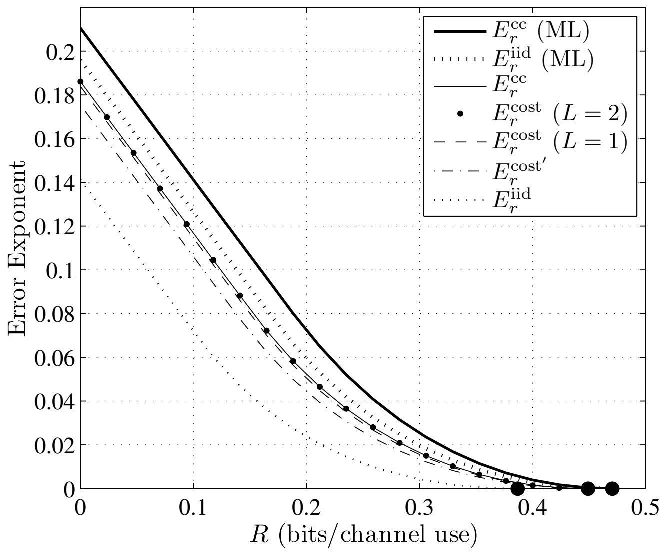

Figure 2. Error exponents for the channel defined in (15) with $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (0.1, 0.3, 0.6)$. The mismatched decoder uses the minimum Hamming distance metric given in (16). The corresponding achievable rates $I^{\mathrm{GMI}}(Q)$, $I^{\mathrm{LM}}(Q)$ and $I(X;Y)$ are respectively marked on the horizontal axis.

the differences between the error exponents when the input distribution is fixed. Under these parameters we have that $I^{\mathrm{GMI}}(Q) = 0.387$, $I^{\mathrm{LM}}(Q) = 0.449$ and $I(X;Y) = 0.471$ bits/use.

We evaluate the exponents using the optimization software YALMIP [20]. The exponent $E_r^{\mathrm{cost}'}$ is optimized over a, performing the optimization jointly with s in (52) and allowing the cost function to vary across rates. The cost function for E_r^{cost} $(L = 1)$ is chosen using an alternating optimization between a and $(s, r, \overline{r})$ in (46), using the a which maximizes $E_0^{\mathrm{cost}'}$ as a starting point and terminating when the change in the exponent between iterations becomes negligible. Similarly, the cost functions for E_r^{cost} $(L = 2)$ are chosen using an alternating optimization between (a_1, a_2) and $(s, r_1, r_2, \overline{r}_1, \overline{r}_2)$, initially setting both a_1 and a_2 to be equal to the a which maximizes $E_0^{\mathrm{cost}'}$.

From Figure 2, we see that E_r^{cc} and E_r^{cost} $(L = 2)$ are indistinguishable, indicating that the alternating optimization technique was effective in finding the true exponent. The exponent E_r^{cost} $(L = 1)$ is only marginally lower, while the gap to $E_r^{\mathrm{cost}'}$ is more significant. The exponent E_r^{iid} is not only lower than each of the other exponents, but also yields a worse achievable rate. This example demonstrates that for a fixed Q, the refined exponent E_r^{cost} $(L = 1)$ can outperform $E_r^{\mathrm{cost}'}$ even when a is optimized.

IV. Conclusion

We have developed a tight characterization of the random-coding error probability for mismatched decoders. A new achievable error exponent has been derived for the cost-constrained i.i.d. ensemble, and alternative forms of existing exponents for the i.i.d. ensemble and constant-composition ensemble have been given. The exponent for each ensemble is tight with respect to the ensemble average for any discrete memoryless channel and decoding metric.

For any given input distribution and cost function, the error exponent for the constant-composition ensemble is at least as high as that of the cost-constrained i.i.d. ensemble, which in turn is at least as high as that of the i.i.d. ensemble. We have shown the error exponent for the constant-composition ensemble can be recovered using cost-constrained i.i.d. random coding with at most two cost constraints.

References

[1] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[2] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU*, vol. 47, no. 4, pp. 228–239, 1993.

[3] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 35–43, Jan. 1995.

[4] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, MIT, 1983.

[5] V. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1889–1902, Nov. 1995.

[6] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[7] A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[8] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[9] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.

[10] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[11] R. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 244–246, March 1973.

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[13] R. Gallager, "Fixed composition arguments and lower bounds to error probability," http://web.mit.edu/gallager/www/notes/notes5.pdf.

[14] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[15] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random-coding bounds," in *Inf. Theory App. Workshop*, La Jolla, CA, 2011.

[16] D. de Caen, "A lower bound on the probability of a union," *Discrete Mathematics*, vol. 169, pp. 217–220, 1997.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[18] K. Fan, "Minimax theorems," *Proc. Nat. Acad. Sci.*, vol. 39, pp. 42–47, 1953.

[19] G. Poltyrev, "Random coding bounds for discrete memoryless channels," *Prob. Inf. Transm.*, vol. 18, no. 1, pp. 9–21, 1982.

[20] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004.