Cost-Constrained Random Coding and Applications

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 a random coding ensemble in which each codeword is constrained to satisfy multiple cost constraints. Using this ensemble, an achievable second-order coding rate is presented for the mismatched single-user channel. Achievable error exponents are given for the mismatched single-user channel and the matched multiple-access channel.**

I. Introduction

The method of random coding is a ubiquitous tool in information theory for proving the existence of codes with a vanishing error probability. In particular, the i.i.d. random-coding ensemble introduced by Shannon [1] has found use in a vast range of settings [2], [3]. An alternative ensemble is the constant-composition ensemble [4], [5], in which each codeword has the same empirical distribution. In many settings, this ensemble yields performance gains over the i.i.d. ensemble. For example, for a given input distribution it is known that the error exponent for the constant-composition ensemble can exceed that of the i.i.d. ensemble [4]. For the multiple-access channel (MAC), the constant-composition ensemble can yield strictly higher error exponents than the i.i.d. ensemble even after the full optimization of the input distributions [6]. Finally, in the setting of mismatched decoding, in which the decoding rule is fixed and only the codebook is optimized, constant-composition random coding not only yields a higher error exponent, but also an improved achievable rate [7].

The cost-constrained i.i.d. ensemble, in which each codeword is randomly generated conditioned on a cost constraint being satisfied, was originally used for settings in which each codeword must satisfy a given system cost (e.g. a power constraint) [8, Ch. 7]. More recently, this ensemble has proved useful as an alternative method for achieving the performance gains of constant-composition codes over i.i.d. codes. For example, the above-mentioned gain in the achievable rate under mismatched decoding can be obtained using the cost-constrained i.i.d. ensemble [9]. In this setting, the cost constraint can be seen as a *pseudo-cost* which is used to improve the performance of the random-coding ensemble. In contrast, system costs are given as part of the problem statement.

In the above applications, the cost-constrained i.i.d. ensemble typically contains only a single cost constraint. In this paper, we consider a similar ensemble with multiple cost constraints. We show that this generalization has many applications in which the performance gains of constant-composition codes can be matched in the discrete memoryless setting, and generalized to infinite and continuous alphabets. In each case, the performance of constant-composition coding is obtained using a fixed number of cost functions which is independent of the alphabet sizes. We present an achievable second-order coding rate under mismatched decoding, and give achievable error exponents for both the mismatched single-user channel and the matched MAC.

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

Notation

The probability of an event is denoted by $\mathbb{P}[\cdot]$. The symbol $\sim$ means "distributed as". The marginals of a joint distribution $P_{XY}(x,y)$ are denoted by $P_X(x)$ and $P_Y(y)$. For a distribution $P_X(x)$, expectations are denoted by $\mathbb{E}_P[\cdot]$, or simply $\mathbb{E}[\cdot]$ when the distribution is understood from the context.

Throughout the paper, we use summations when writing expectations explicitly (e.g. $\mathbb{E}[g(X)] = \sum_x P_X(x)g(x)$). However, it should be noted that the alphabets are not assumed to be finite, and the results apply to continuous alphabets when the summations are replaced by integrals.

Given a distribution $Q(x)$ and a conditional distribution $W(y|x)$, we write $Q \times W$ to denote the joint distribution $Q(x)W(y|x)$. Mutual information with respect to a joint distribution $P_{XY}(x,y)$ is written as $I_P(X;Y)$. All logarithms have base e, and all rates are in units of nats. We denote the indicator function by $\mathbb{1}\{\cdot\}$.

For two functions $f(n)$ and $g(n)$, we write $f(n) = O(g(n))$ if $f(n) \le cg(n)$ for some constant c and for sufficiently large n. We write $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$, and $f(n) = o(g(n))$ if $\frac{f(n)}{g(n)} \to 0$.

II. Random-Coding Ensemble

In this paper, we are interested in setups involving block coding, in which an encoder selects a message m equiprobably from the set $\{1,\dots,M\}$ and transmits the corresponding codeword $\boldsymbol{x}^{(m)}$ from a codebook $\mathcal{C} = \{\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(M)}\}$. We consider random coding, in which each codeword is independently generated according to a distribution $P_{\boldsymbol{X}}(\boldsymbol{x})$.

The cost-constrained i.i.d. ensemble with L cost constraints is given by

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\mu_n}\prod_{i=1}^{n} Q(x_i)\mathbb{1}\big\{\boldsymbol{x} \in \mathcal{D}_n\big\}, \tag{1}$$

where

$$\mathcal{D}_n \triangleq \left\{ \boldsymbol{x} : \left| \frac{1}{n} \sum_{i=1}^{n} a_l(x_i) - \phi_l \right| \leq \frac{\delta}{n},\, l = 1, \dots, L \right\}, \quad (2)$$

and where μ_n is a normalizing constant, δ is a positive constant (independent of n), and for each $l \in \{1, \dots, L\}$, $a_l(\cdot)$ is a cost function and $\phi_l \triangleq \mathbb{E}_Q[a_l(X)]$. Roughly speaking, each codeword is generated according to an i.i.d. distribution conditioned on $\frac{1}{n}\sum_i a_l(x_i)$ being close to its mean for all l.

The ensemble described by (1)–(2) can be generalized in several ways. The constant δ could vary with both n and l, but a fixed value will suffice for our purposes. More generally, one could consider constraints of the form

$$\frac{\delta'_{l,n}}{n} \leq \frac{1}{n} \sum_{i=1}^{n} a_l(x_i) - \phi_l \leq \frac{\delta''_{l,n}}{n}. \quad (3)$$

In particular, the choice $\delta''_{l,n} = 0$ is relevant in the case that each codeword is constrained to satisfy a system cost [8]. The results of this paper can easily be extended to this setting. However, our focus is on the improvements obtained by introducing pseudo-costs, and thus we do not include system costs in the problem formulation.

If $|\mathcal{X}|$ is finite, then the constant-composition ensemble is a special case of the above ensemble, since it is obtained by setting $L = |\mathcal{X}| - 1$ and $\delta < 1$ for all l, and choosing the cost functions $a_1 = (1, 0, \dots, 0)$, $a_2 = (0, 1, 0, \dots, 0)$, etc. However, we will see that many of the performance gains of constant-composition codes can be obtained using a fixed number of cost functions which is independent of $|\mathcal{X}|$.

The following proposition shows that μ_n decays polynomially in n. This result will prove to be important for the applications presented in the later sections. A proof for the case $L = 1$ is given in [8], [10].

Proposition 1. *If the second moment of $a_l(X)$ is finite under $X \sim Q$ for all $l = 1, \dots, L$, then $\mu_n = \Omega(n^{-L/2})$.*

Proof: We extend the steps of [10, Eq. (88)] to the multidimensional setting. Defining the vectors $\boldsymbol{a}(x) = (a_1(x), \dots, a_L(x))^T$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_L)^T$, we have

$$\mu_n = \mathbb{P}\left[-\frac{\delta}{\sqrt{n}}\mathbf{1} \preceq \frac{1}{\sqrt{n}}\big(\boldsymbol{a}(X) - \boldsymbol{\phi}\big) \preceq \frac{\delta}{\sqrt{n}}\mathbf{1} \right], \quad (4)$$

where $\preceq$ denotes element-wise inequality and $\mathbf{1}$ is the vector of ones. We define Σ to be the covariance matrix of $\boldsymbol{a}(X) - \boldsymbol{\phi}$. We can assume without loss of generality that $\det(\Sigma) > 0$, since otherwise we can reduce the evaluation of μ_n to a lower dimension.[1] Since each ϕ_l and $\sigma_l^2 \triangleq \mathbb{E}[(a_l(X) - \phi_l)^2]$ is finite by assumption, we can apply the multidimensional central limit theorem [11], yielding

$$\mu_n \to \mathbb{P}\left[-\frac{\delta}{\sqrt{n}}\mathbf{1} \preceq \boldsymbol{Z} \preceq \frac{\delta}{\sqrt{n}}\mathbf{1} \right], \quad (5)$$

where $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ has density

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{L/2}\det(\Sigma)^{1/2}} \exp\Big(-\frac{1}{2}\boldsymbol{z}^T \Sigma^{-1} \boldsymbol{z} \Big). \quad (6)$$

To approximate (5), we use the fact that $f_{\boldsymbol{Z}}(\boldsymbol{z})$ approaches $\frac{1}{(2\pi)^{L/2}\det(\Sigma)^{1/2}}$ as each entry of $\boldsymbol{z}$ tends to zero, and hence

$$\mu_n \to \frac{\left(\frac{2\delta}{\sqrt{n}}\right)^L}{(2\pi)^{L/2}\det(\Sigma)^{1/2}} \quad (7)$$

since $\left(\frac{2\delta}{\sqrt{n}}\right)^L$ is the area of the region over which the integration is being performed. Equations (5) and (7) can be made more precise by using the Berry-Esseen theorem [11], yielding the growth of $\Omega(n^{-L/2})$ in (7). ■

III. Second-Order Coding Rate for Mismatched Decoding

In this section, we consider the problem of channel coding with a given decoding rule (mismatched decoding) [7], [9]. We consider block coding, as described at the beginning of Section II. Upon receiving the signal $\boldsymbol{y}$ at the output of the memoryless channel $W(y|x)$, the decoder forms the estimate[2]

$$\hat{m} = \arg\max_j \prod_{i=1}^{n} q(x_i^{(j)}, y_i), \quad (8)$$

where $q(x, y)$ is a given (possibly suboptimal) non-negative decoding metric, and $x_i^{(j)}$ (respectively, y_i) denotes the i-th entry of $\boldsymbol{x}^{(j)}$ (respectively, $\boldsymbol{y}$). An error is said to have occurred if $\hat{m} \neq m$. We say that (M, n, ϵ) is achievable if there exists a code with M codewords of block length n with average error probability under the decoding metric $q(x, y)$ not exceeding ϵ. For a given error probability ϵ and block length n, the largest M such that (M, n, ϵ) is achievable is denoted by $M^*(n, \epsilon)$. We say that R is an achievable rate if

$$\log M^*(n, \epsilon) \geq nR - o(n)$$

for all $\epsilon \in (0, 1)$.

For memoryless channels, cost-constrained coding can be used to prove the achievability of the LM rate [9], given by

$$I^{\mathrm{LM}}(Q) \triangleq \sup_{s \geq 0, a(\cdot)} \mathbb{E}\left[\log \frac{q(X,Y)^s e^{a(X)}}{\mathbb{E}[q(\overline{X}, Y)^s e^{a(\overline{X})} \,|\, Y]} \right], \quad (9)$$

where Q is an arbitrary input distribution. In this section, we show that multiple cost constraints can be used to obtain a corresponding achievable second-order coding rate; this statement is made more precise in Theorem 1 below.

We fix an input distribution $Q(x)$ and define the quantities

$$i_{s,a}(x, y) \triangleq \log \frac{q(x,y)^s e^{a(x)}}{\mathbb{E}[q(\overline{X}, y)^s e^{a(\overline{X})}]} \quad (10)$$

$$I_{s,a}(Q) \triangleq \mathbb{E}[i_{s,a}(X, Y)] \quad (11)$$

[1] This statement may not be true in general under constraints of the form (3). A simple counter-example is given by $L = 2$, $\delta''_{1,n} = \delta''_{2,n} = 0$, and $a_1(x) = -a_2(x)$, yielding an overall constraint of $\frac{1}{n}\sum_{i=1}^{n} a_1(x_i) = \phi_1$.

[2] Our analysis is unaffected under any method for breaking ties, since we upper bound the error probability by that of the decoder which decodes ties as errors.

$$V_{s,a}(Q) \triangleq \mathbb{E}\big[\mathrm{Var}[i_{s,a}(X,Y)\,|\,X]\big], \quad (12)$$

where $(X,Y,\overline{X}) \sim Q(x)W(y|x)Q(\overline{x})$. The quantity $i_{s,a}$ can be viewed as a generalized information density, since the information density [12] is recovered by substituting $s=1$, $a(x)=0$ and $q(x,y)=W(y|x)$. We observe that $I_{s,a}(Q)$ coincides with $I^{\mathrm{LM}}(Q)$ after the optimization of s and a.

We further define

$$i_{s,a}(x) \triangleq \mathbb{E}[i_{s,a}(x,Y)] \quad (13)$$

$$v_{s,a}(x) \triangleq \mathrm{Var}[i_{s,a}(x,Y)] \quad (14)$$

$$t_{s,a}(x) \triangleq \mathbb{E}[|i_{s,a}(x,Y) - I_{s,a}(Q)|^3], \quad (15)$$

where each expectation is taken with respect to $W(\cdot|x)$. We write $i_{s,a}(\boldsymbol{x},\boldsymbol{y})$ as a shorthand for $\sum_{i=1}^n i_{s,a}(x_i,y_i)$, and similarly for $a_l(\boldsymbol{x})$, $v_{s,a}(\boldsymbol{x})$ and $t_{s,a}(\boldsymbol{x})$. Similarly, we write $q(\boldsymbol{x},\boldsymbol{y})$, $W(\boldsymbol{y}|\boldsymbol{x})$ and $Q(\boldsymbol{x})$ to denote the multiplicative extension of the single-letter definition (e.g. $q(\boldsymbol{x},\boldsymbol{y}) = \prod_{i=1}^n q(x_i,y_i)$). We define the random variables

$$(\boldsymbol{X},\boldsymbol{Y},\overline{\boldsymbol{X}}) \sim P_{\boldsymbol{X}}(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x})P_{\boldsymbol{X}}(\overline{\boldsymbol{x}}) \quad (16)$$

The following assumption on Q, W, s and a will be used.

Assumption 1. Let $\boldsymbol{X}' \sim \prod_{i=1}^n Q(x_i')$. For any $\delta_v > 0$ (independent of n), $v_{s,a}$ satisfies

$$\mathbb{P}\Big[\frac{1}{n}v_{s,a}(\boldsymbol{X}') - V_{s,a}(Q) \ge \delta_v\Big] = o\Big(\frac{1}{n}\Big) \quad (17)$$

and for some $\delta_t < \infty$ (independent of n), $t_{s,a}$ satisfies

$$\mathbb{P}\Big[\frac{1}{n}t_{s,a}(\boldsymbol{X}') - \mathbb{E}_Q[t_{s,a}(X)] \ge \delta_t\Big] = o\Big(\frac{1}{n}\Big). \quad (18)$$

The right-hand sides of (17)–(18) are written as $o\big(\frac{1}{n}\big)$ for the sake of generality. To fulfill the assumption, it clearly suffices for the left-hand sides of (17)–(18) to be $O(e^{-\psi n})$ for some constant ψ. By Hoeffding's inequality [11], this occurs when $v_{s,a}(x)$ and $t_{s,a}(x)$ are bounded, which in turn occurs in the discrete memoryless setting. From [13, Ex. 2.1.6], a more general sufficient condition is that $v_{s,a}(X)$ and $t_{s,a}(X)$ are sub-Gaussian under $X \sim Q$, i.e. $\mathbb{P}[v_{s,a}(X) \ge v] \le Ae^{-cv^2}$ for some constants A and c, and similarly for $t_{s,a}$.

In the following theorem, $\mathsf{Q}^{-1}(\cdot)$ denotes its functional inverse of $\mathsf{Q}(\cdot)$, defined to be the upper tail probability of a zero-mean unit-variance Gaussian random variable.

Theorem 1. *For any $s > 0$, $a(x)$ and distribution $Q(x)$, $M^*(n,\epsilon)$ satisfies*

$$\log M^*(n,\epsilon) \ge nI_{s,a}(Q) - \sqrt{nV_{s,a}(Q)}\,\mathsf{Q}^{-1}(\epsilon) + o(\sqrt{n}) \quad (19)$$

provided that, under $(X,Y) \sim Q \times W$, the second moment of $a(X)$ is finite, the third absolute moment of $i_{s,a}(X,Y)$ is finite, and Assumption 1 holds.

Proof: We set $L=2$ and choose the cost functions

$$a_1(x) = a(x) \quad (20)$$

$$a_2(x) = i_{s,a}(x). \quad (21)$$

By the assumptions on the moments of $a(x)$ and $i_{s,a}(x,y)$, and using Proposition 1, we obtain $\mu_n = \Omega(n^{-1})$.

The random-coding union (RCU) bound for mismatched decoders [14], [15] can be written as

$$\begin{aligned}
\overline{p}_e &\le \mathbb{E}\Big[\min\Big\{1,(M-1) \\
&\quad\times \mathbb{P}[i_{s,a}(\overline{\boldsymbol{X}},\boldsymbol{Y}) - a(\overline{\boldsymbol{X}}) \ge i_{s,a}(\boldsymbol{X},\boldsymbol{Y}) - a(\boldsymbol{X})\,|\,\boldsymbol{X},\boldsymbol{Y}]\Big\}\Big] && (22)\\
&\le \mathbb{E}\Big[\min\Big\{1,(M-1) \\
&\quad\times \mathbb{P}[i_{s,a}(\overline{\boldsymbol{X}},\boldsymbol{Y}) \ge i_{s,a}(\boldsymbol{X},\boldsymbol{Y}) - 2\delta\,|\,\boldsymbol{X},\boldsymbol{Y}]\Big\}\Big] && (23)\\
&\le \mathbb{P}\big[i_{s,a}(\boldsymbol{X},\boldsymbol{Y}) \le \gamma + 2\delta\big] + (M-1)\mathbb{P}\big[i_{s,a}(\overline{\boldsymbol{X}},\boldsymbol{Y}) > \gamma\big], && (24)
\end{aligned}$$

where (23) follows from the bounds on $a(\boldsymbol{x})$ and $a(\overline{\boldsymbol{x}})$ given in the constraint set in (2), and (24) follows by upper bounding the minimum in (22) by 1 when $i_{s,a}(\boldsymbol{X},\boldsymbol{Y}) \le \gamma + 2\delta$, and by the second term otherwise. The first probability in (24) can be upper bounded as

$$\begin{aligned}
&\mathbb{P}\big[i_{s,a}(\boldsymbol{X},\boldsymbol{Y}) \le \gamma + 2\delta\big] && (25)\\
&\quad\le \mathbb{P}\big[\boldsymbol{X} \notin \mathcal{A}_n\big] + \max_{\boldsymbol{x} \in \mathcal{A}_n} \mathbb{P}\big[i_{s,a}(\boldsymbol{x},\boldsymbol{Y}) \le \gamma + 2\delta\big], && (26)
\end{aligned}$$

where the set $\mathcal{A}_n$ is arbitrary. We fix $\delta_v > 0$ and $\delta_t < \infty$ (the latter satisfying (18) in Assumption 1) and choose

$$\mathcal{A}_n \triangleq \Big\{\boldsymbol{x} \in \mathcal{D}_n \,:\, v_{s,a}(\boldsymbol{x}) \le n\big(V_{s,a}(Q) + \delta_v\big), \; t_{s,a}(\boldsymbol{x}) \le n\big(\mathbb{E}_Q[t_{s,a}(X)] + \delta_t\big)\Big\}. \quad (27)$$

We thus have from the union bound that

$$\begin{aligned}
\mathbb{P}\big[\boldsymbol{X} \notin \mathcal{A}_n\big] &\le \mathbb{P}\Big[v_{s,a}(\boldsymbol{X}) > n\big(V_{s,a}(Q) + \delta_v\big)\Big] && (28)\\
&\quad + \mathbb{P}\Big[t_{s,a}(\boldsymbol{X}) > n\big(\mathbb{E}_Q[t_{s,a}(X)] + \delta_t\big)\Big] \\
&\le \frac{1}{\mu_n}\mathbb{P}\Big[v_{s,a}(\boldsymbol{X}') > n\big(V_{s,a}(Q) + \delta_v\big)\Big] && (29)\\
&\quad + \frac{1}{\mu_n}\mathbb{P}\Big[t_{s,a}(\boldsymbol{X}') > n\big(\mathbb{E}_Q[t_{s,a}(X)] + \delta_t\big)\Big] \\
&= o(1), && (30)
\end{aligned}$$

where (29) follows from (1) and by defining $\boldsymbol{X}' \sim \prod_{i=1}^n Q(x_i')$, and (30) follows since $\mu_n = \Omega(n^{-1})$ and by Assumption 1. Combining (26) and (30), we obtain

$$\mathbb{P}\big[i_{s,a}(\boldsymbol{X},\boldsymbol{Y}) \le \gamma + 2\delta\big] \le \max_{\boldsymbol{x} \in \mathcal{A}_n} \mathbb{P}\big[i_{s,a}(\boldsymbol{x},\boldsymbol{Y}) \le \gamma + 2\delta\big] + o(1). \quad (31)$$

The second probability in (24) can be written as

$$\begin{aligned}
&\sum_{\overline{\boldsymbol{x}},\boldsymbol{y}} P_{\boldsymbol{X}}(\overline{\boldsymbol{x}})P_{\boldsymbol{Y}}(\boldsymbol{y})\mathbf{1}\{i_{s,a}(\overline{\boldsymbol{x}},\boldsymbol{y}) > \gamma\} && (32)\\
&\le \frac{1}{\mu_n}\sum_{\overline{\boldsymbol{x}},\boldsymbol{y}} Q(\overline{\boldsymbol{x}})P_{\boldsymbol{Y}}(\boldsymbol{y})\mathbf{1}\{i_{s,a}(\overline{\boldsymbol{x}},\boldsymbol{y}) > \gamma\} && (33)\\
&\le \frac{1}{\mu_n}\sum_{\overline{\boldsymbol{x}},\boldsymbol{y}} Q(\overline{\boldsymbol{x}})P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{q(\overline{\boldsymbol{x}},\boldsymbol{y})^s e^{a(\overline{\boldsymbol{x}})}}{\sum_{\boldsymbol{x}'} Q(\boldsymbol{x}')q(\boldsymbol{x}',\boldsymbol{y})^s e^{a(\boldsymbol{x}')}}e^{-\gamma} && (34)\\
&= \frac{1}{\mu_n}e^{-\gamma}, && (35)
\end{aligned}$$

where (33) follows by substituting the definition of the random-coding ensemble and by summing over all $\overline{\boldsymbol{x}}$ (rather than just $\overline{\boldsymbol{x}} \in \mathcal{D}_n$), and (34) follows by using the definition of $i_{s,a}$ and upper bounding the indicator function.

In order to apply the Berry-Esseen theorem [11] to (31), we must bound the relevant first, second and third moments associated with $i_{s,a}(\boldsymbol{x}, \boldsymbol{Y})$ for $\boldsymbol{x} \in \mathcal{A}_n$. The second and third moments are already bounded by the definition of $\mathcal{A}_n$ in (27); recall also that the third absolute moment of $i_{s,a}(X,Y)$ is finite by assumption. For the first moment, we have

$$\mathbb{E}[i_{s,a}(\boldsymbol{x}, \boldsymbol{Y})] = nI_{s,a}(Q) + O(1) \tag{36}$$

for all $\boldsymbol{x} \in \mathcal{D}_n$, which follows from the choice of a_2 in (21) and the definitions of $i_{s,a}(x)$ and $\mathcal{D}_n$. Applying the Berry-Esseen theorem accordingly and using an identical argument to [16, Theorem 1], we obtain from (24), (31) and (35) that

$$\log M^*(n,\epsilon) \ge nI_{s,a}(Q) - \sqrt{n\big(V_{s,a}(Q) + \delta_v\big)}\mathsf{Q}^{-1}(\epsilon) + o(\sqrt{n}). \tag{37}$$

The proof follows by taking $\delta_v \to 0$ and applying a first-order Taylor expansion to the square root. ∎

In the discrete memoryless setting, the expansion in (19) can be obtained using constant-composition random coding by following the analysis of [16]. The proof of Theorem 1 shows that the same expansion is obtained using cost-constrained coding with $L = 2$.

Under suitable technical assumptions, the term $o(\sqrt{n})$ in (19) can be improved to $O(\log n)$. If the assumption in (17) is replaced by the assumption that

$$\mathbb{P}\Big[\frac{1}{n}v_{s,a}(\boldsymbol{X}') - V_{s,a}(Q) \ge \delta_{v,n}\Big] = O\Big(\frac{\log n}{n\sqrt{n}}\Big) \tag{38}$$

for some $\delta_{v,n} = O\big(\frac{\log n}{\sqrt{n}}\big)$, then this improvement is obtained using identical steps to Theorem 1. Alternatively, one can set $L = 3$ and let $a_3(x) = v_{s,a}(x)$, thus ensuring that $\mathrm{Var}[i_{s,a}(\boldsymbol{x}, \boldsymbol{Y})]$ is close to its mean, similarly to $\mathbb{E}[i_{s,a}(\boldsymbol{x}, \boldsymbol{Y})]$ in (36). In this case, Proposition 1 requires that the variance of $v_{s,a}(X)$ is finite, which follows provided that the fourth moment of $i_{s,a}(X,Y)$ is finite. We no longer require (17) to hold, but the right-hand side of (18) must be replaced by $O\big(\frac{\log n}{n^2}\big)$ in order to achieve the third-order $O(\log n)$ term.

Assuming the supremum is achieved in (9), the best asymptotic expansion in (19) for a given Q is obtained by letting (s,a) be a pair which maximizes $I_{s,a}(Q)$. If multiple such (s,a) exist and $\epsilon < \frac{1}{2}$, the best expansion is obtained by choosing the one which minimizes $V_{s,a}(Q)$. In any case, one must ensure that the resulting (s,a) are such that the assumptions of Theorem 1 are satisfied. In the discrete memoryless setting, this is always the case.

IV. Error Exponent for Mismatched Decoding

In this section, we use the cost-constrained ensemble with multiple constraints to obtain an achievable error exponent for mismatched decoding. More formally, we consider the same setting as Section III, and we say that $E(R)$ is an achievable error exponent if there exists a sequence of codebooks of length n and rate R whose error probability satisfies

$$\lim_{n\to\infty} -\frac{1}{n}\log p_e \ge E(R) \tag{39}$$

under the decoding metric $q(x,y)$. We henceforth write $f(\{a_l\})$ to denote a function f which depends on the cost functions $a_1, \dots, a_L$.

Theorem 2. *Fix any distribution Q and cost functions $\{a_l\}_{l=1}^L$. If the second moment of $a_l(X)$ is finite for all $l = 1, \dots, L$, then an achievable error exponent is given by*

$$E_r^{\mathrm{cost}}(Q, R, \{a_l\}) = \max_{\rho\in[0,1]} E_0^{\mathrm{cost}}(Q,\rho,\{a_l\}) - \rho R, \tag{40}$$

where

$$E_0^{\mathrm{cost}}(Q,\rho,\{a_l\}) \triangleq \sup_{s\ge 0, \{r_l\}, \{\overline{r}_l\}} -\log\mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X},Y)^s e^{\sum_{l=1}^L \overline{r}_l(a_l(\overline{X})-\phi_l)} \,|\, Y\big]}{q(X,Y)^s e^{\sum_{l=1}^L r_l(a_l(X)-\phi_l)}}\right)^\rho\right]. \tag{41}$$

Proof: Applying Markov's inequality and $\min\{1,\alpha\} \le \alpha^\rho$ ($\rho \in [0,1]$) to (22), we obtain

$$\overline{p}_e \le \frac{1}{\mu_n^{1+\rho}}M^\rho \sum_{\boldsymbol{x}\in\mathcal{D}_n, \boldsymbol{y}} Q(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x}) \times \left(\frac{\sum_{\overline{\boldsymbol{x}}\in\mathcal{D}_n} Q(\overline{\boldsymbol{x}})q(\overline{\boldsymbol{x}},\boldsymbol{y})^s}{q(\boldsymbol{x},\boldsymbol{y})^s}\right)^\rho \tag{42}$$

for any $\rho \in [0,1]$ and $s \ge 0$. It follows from (2) that each codeword $\boldsymbol{x} \in \mathcal{D}_n$ satisfies

$$\exp\big(r_l(a_l(\boldsymbol{x}) - n\phi_l)\big) \le e^{|r_l|\delta} \tag{43}$$

for any real number r_l. Thus, combining (42) with (43), we obtain

$$\overline{p}_e \le \frac{e^{\rho\sum_l(|r_l|+|\overline{r}_l|)\delta}}{\mu_n^{1+\rho}}M^\rho \sum_{\boldsymbol{x}} Q(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x}) \times \left(\frac{\sum_{\overline{\boldsymbol{x}}} Q(\overline{\boldsymbol{x}})q(\overline{\boldsymbol{x}},\boldsymbol{y})^s e^{\sum_l \overline{r}_l(a_l(\overline{\boldsymbol{x}})-n\phi_l)}}{q(\boldsymbol{x},\boldsymbol{y})^s e^{\sum_l r_l(a_l(\boldsymbol{x})-n\phi_l)}}\right)^\rho \tag{44}$$

for any real numbers $\{r_l\}$ and $\{\overline{r}_l\}$, where we have replaced the summations over $\mathcal{D}_n$ with summations over all sequences. Expanding each term in the summation as product from 1 to n, and noting from Proposition 1 that μ_n decays to zero subexponentially in n, we obtain the exponent in (40)–(41). ∎

In [15], the present authors derived E_r^{cost} in the discrete memoryless setting using the method of types and Lagrange duality. Connections were drawn between E_r^{cost} and the exponent E_r^{cc} for constant-composition random coding, given by

$$E_r^{\mathrm{cc}}(Q,R) \triangleq \max_{\rho\in[0,1]} E_0^{\mathrm{cc}}(Q,\rho) - \rho R \tag{45}$$

$$E_0^{\rm cc}(Q,\rho) \triangleq \sup_{s\ge 0, a(\cdot)} \mathbb{E}\left[-\log \mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X},Y)^s e^{a(\overline{X})}\,|\,Y\big]}{q(X,Y)^s e^{a(X)}}\right)^{\rho}\,\Bigg|\, X\right]\right]. \quad (46)$$

In particular, it was shown that $E_r^{\rm cc} \ge E_r^{\rm cost}$, with equality when $L=2$ suitably chosen cost functions are used. In the remainder of this section, we give a direct derivation of $E_r^{\rm cc}$ using cost-constrained coding, again using $L=2$. As was done in [15], we let one of the cost functions play the role of $a(x)$ in (46). In contrast to [15], we also give an explicit formula for the second cost function in terms of the first.

Fix the input distribution Q, and the parameters $\rho\in[0,1]$, $s\ge 0$ and $a(x)$. We set $a_1(x)=a(x)$ and

$$a_2(x) = \log\sum_y W(y|x)\bigg(\sum_{\overline{x}} Q(\overline{x})\Big(\frac{q(\overline{x},y)}{q(x,y)}\Big)^s\frac{e^{a(\overline{x})}}{e^{a(x)}}\bigg)^{\rho}. \quad (47)$$

Combining (42) and (43), $\overline{p}_e$ is upper bounded by

$$\frac{e^{2\rho\delta}}{\mu_n^{\rho}}\sum_{\boldsymbol{x},\boldsymbol{y}} P_{\boldsymbol{X}}(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x})\bigg(M\sum_{\overline{\boldsymbol{x}}}Q(\overline{\boldsymbol{x}})\Big(\frac{q(\overline{\boldsymbol{x}},\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\Big)^s\frac{e^{a(\overline{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}}\bigg)^{\rho} \quad (48)$$

$$\le \frac{e^{2\rho\delta}}{\mu_n^{\rho}}M^{\rho}\max_{\boldsymbol{x}\in\mathcal{D}_n}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x})\bigg(\sum_{\overline{\boldsymbol{x}}}Q(\overline{\boldsymbol{x}})\Big(\frac{q(\overline{\boldsymbol{x}},\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\Big)^s\frac{e^{a(\overline{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}}\bigg)^{\rho}. \quad (49)$$

We write the summation in (49) as

$$\exp\Bigg(\log\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x})\bigg(\sum_{\overline{\boldsymbol{x}}}Q(\overline{\boldsymbol{x}})\Big(\frac{q(\overline{\boldsymbol{x}},\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\Big)^s\frac{e^{a(\overline{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}}\bigg)^{\rho}\Bigg). \quad (50)$$

The constraint on $a_2(x)$ in the definition of $\mathcal{D}_n$ implies that, for all $\boldsymbol{x}\in\mathcal{D}_n$, the logarithm in (50) differs from its mean by no more than δ. Expanding each term in the logarithm as a product from 1 to n, we see that this mean is simply $nE_0^{\rm cc}$, and thus (50) is upper bounded by

$$\exp\big(nE_0^{\rm cc}(Q,\rho)+\delta\big). \quad (51)$$

The derivation is concluded by substituting (51) into (49) and using Proposition 1, and maximizing over $\rho\in[0,1]$, $s\ge 0$ and a. In the case of infinite or continuous alphabets, the supremum in (46) is restricted to parameters such that the resulting second moments of $a_1(X)=a(X)$ and $a_2(X)$ are finite under $X\sim Q$, in accordance with Proposition 1.

V. Error Exponent for the Multiple-Access Channel

In this section, we consider the MAC with two users. Given the codebooks $\mathcal{C}_\nu = \{\boldsymbol{x}_\nu^{(1)},\dots,\boldsymbol{x}_\nu^{(M)}\}$ ($\nu=1,2$), user ν transmits the codeword $\boldsymbol{x}_\nu^{(m_\nu)}$ corresponding to a randomly chosen message m_ν. Upon receiving the output signal $\boldsymbol{y}$ at the output of the (memoryless) channel $W(y|x_1,x_2)$, the decoder forms the estimate $(\hat{m}_1,\hat{m}_2)$. We limit our attention to the unconstrained memoryless MAC with maximum-likelihood (ML) decoding. Extensions to the constrained and mismatched settings are possible, though the analysis in the latter setting is somewhat more involved (e.g. see [17]).

We use similar terminology to the single-user setting for the achievability of rate pairs and error exponents (e.g. see [18]). We focus on random coding in the absence of time-sharing, and outline the corresponding results with time-sharing in Remark 1. Thus, we consider random coding of the form

$$P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu) = \frac{1}{\mu_{\nu,n}}\prod_{i=1}^n Q_\nu(x_{\nu,i})\mathbb{1}\big\{\boldsymbol{x}_\nu\in\mathcal{D}_{\nu,n}\big\}, \quad (52)$$

where

$$\mathcal{D}_{\nu,n} \triangleq \left\{\boldsymbol{x}_\nu \,:\, \left|\frac{1}{n}\sum_{i=1}^n a_{\nu,l}(x_{\nu,i})-\phi_{\nu,l}\right| \le \frac{\delta}{n},\, l=1,\dots,L_\nu\right\} \quad (53)$$

for $\nu=1,2$, and where each quantity is defined analogously to (1)–(2). For user ν, we randomly and independently generate $M_\nu = e^{nR_\nu}$ codewords according to $P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu)$.

We split the error event into three types:

(Type 1) $\hat{m}_1\ne m_1$ and $\hat{m}_2 = m_2$
(Type 2) $\hat{m}_1 = m_1$ and $\hat{m}_2 \ne m_2$
(Type 12) $\hat{m}_1\ne m_1$ and $\hat{m}_2 \ne m_2$

The random-coding error probabilities of these events are denoted by $\overline{p}_{e,1}$, $\overline{p}_{e,2}$ and $\overline{p}_{e,12}$ respectively. To ease the notation, we write $f(\boldsymbol{Q})$ to denote a function which depends on the input distributions Q_1 and Q_2.

The following theorem states an achievable random-coding error exponent for the above ensemble. In the proof, we first obtain an error exponent for a general choice of L_1 and L_2, and then show that it is maximized by choosing $L_1=L_2=3$ and optimizing the corresponding cost functions.

Theorem 3. *For any (Q_1,Q_2), an achievable error exponent for the memoryless MAC is given by*

$$E_r(\boldsymbol{Q},R_1,R_2) \triangleq \min\Big\{E_{r,1}(\boldsymbol{Q},R_1), E_{r,2}(\boldsymbol{Q},R_2), E_{r,12}(\boldsymbol{Q},R_1,R_2)\Big\}, \quad (54)$$

where

$$E_{r,1}(\boldsymbol{Q},R_1) \triangleq \sup_{\rho\in[0,1]} E_{0,1}(\boldsymbol{Q},\rho)-\rho R_1 \quad (55)$$

$$E_{r,2}(\boldsymbol{Q},R_2) \triangleq \sup_{\rho\in[0,1]} E_{0,2}(\boldsymbol{Q},\rho)-\rho R_2 \quad (56)$$

$$E_{r,12}(\boldsymbol{Q},R_1,R_2) \triangleq \sup_{\rho\in[0,1]} E_{0,12}(\boldsymbol{Q},\rho)-\rho(R_1+R_2) \quad (57)$$

$$E_{0,1}(\boldsymbol{Q},\rho) \triangleq \sup_{a_1(\cdot),a_2(\cdot)} -\log\sum_{x_2,y}Q_2(x_2) \times\bigg(\sum_{x_1}Q_1(x_1)W(y|x_1,x_2)^{\frac{1}{1+\rho}}e^{\sum_{\nu=1}^2 a_\nu(x_\nu)-\phi_\nu}\bigg)^{1+\rho} \quad (58)$$

$$E_{0,2}(\boldsymbol{Q},\rho) \triangleq \sup_{a_1(\cdot),a_2(\cdot)} -\log \sum_{x_1,y} Q_1(x_1) \times \Bigg(\sum_{x_2} Q_2(x_2) W(y|x_1,x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^{2} a_\nu(x_\nu)-\phi_\nu} \Bigg)^{1+\rho} \quad (59)$$

$$E_{0,12}(\boldsymbol{Q},\rho) \triangleq \sup_{a_1(\cdot),a_2(\cdot)} -\log \sum_{y} \Bigg(\sum_{x_1,x_2} Q_1(x_1)Q_2(x_2) \times W(y|x_1,x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^{2} a_\nu(x_\nu)-\phi_\nu} \Bigg)^{1+\rho}. \quad (60)$$

Each supremum over a_ν is taken over all functions such that the second moment of $a_\nu(X_\nu)$ is finite under $X_\nu \sim Q_\nu$.

Proof: We first analyze the cost-constrained i.i.d. ensemble for a given L_1 and L_2. Each error event is handled similarly, so we focus on the type-12 event. We analyze a straightforward extension of the RCU bound [19], given by

$$\overline{p}_{e,12} \le \mathbb{E}\Big[\min\Big\{1, (M_1-1)(M_2-1) \times \mathbb{P}\big[W(\boldsymbol{Y}|\overline{\boldsymbol{X}}_1,\overline{\boldsymbol{X}}_2) \ge W(\boldsymbol{Y}|\boldsymbol{X}_1,\boldsymbol{X}_2) \,|\, \boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}\big]\Big\}\Big]. \quad (61)$$

Applying analogous steps to the proof of Theorem 2, we obtain the E_0 function

$$E_{0,12}(\boldsymbol{Q},\rho,\{a_l\}) \triangleq -\log \mathbb{E}\left[\left(\frac{\mathbb{E}\big[W(Y|\overline{X}_1,\overline{X}_2)^s e^{\sum_{\nu=1}^{2}\sum_{l=1}^{L_\nu} \overline{r}_{\nu,l}\left(a_{\nu,l}(x_\nu)-\phi_{\nu,l}\right)} \,\big|\, Y\big]}{W(Y|X_1,X_2)^s e^{\sum_{\nu=1}^{2}\sum_{l=1}^{L_\nu} r_{\nu,l}\left(a_{\nu,l}(x_\nu)-\phi_{\nu,l}\right)}}\right)^{\rho}\right] \quad (62)$$

for arbitrary constants $\{r_{\nu,l}\}$ and $\{\overline{r}_{\nu,l}\}$. Using Hölder's inequality similarly to [8, Ex. 5.6], we can lower bound the expectation in (62) by

$$\sum_{y} \Bigg(\sum_{x_1,x_2} Q_1(x_1)Q_2(x_2) \times W(y|x_1,x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^{2}\sum_{l=1}^{L_\nu} \overline{r}_{\nu,l}\left(a_{\nu,l}(x_\nu)-\phi_{\nu,l}\right)} \Bigg)^{1+\rho} \quad (63)$$

and obtain the parameters $s=\frac{1}{1+\rho}$ and $r_{\nu,l}=\frac{-\overline{r}_{\nu,l}}{\rho}$ (for all ν,l) achieving this bound. Similar E_0 functions are obtained for the type-1 and type-2 events.

To obtain (58)–(60), we set $L_1=L_2=3$ and choose $\overline{r}_{\nu,l}$ as follows. For $\nu=1,2$, we let $\overline{r}_{\nu,l}=1$ for one value of l, and $\overline{r}_{\nu,l}=0$ for the other two values of l. Since there are three error events and three cost functions per user, a different cost function can be used for each error event. Using this observation and optimizing each cost function accordingly, we obtain (58)–(60). We can do no better by introducing further cost functions, since for $\nu=1,2$ any sum of cost functions $a_{\nu,l}(x_\nu)$ weighted by $\overline{r}_{\nu,l}$ in (63) can be replaced by an equivalent single cost function $\sum_l \overline{r}_{\nu,l} a_{\nu,l}(x_\nu)$. ■

Remark 1. It is well known that time-sharing can improve both the achievable rates and error exponents of the MAC [6], [18]. The cost-constrained ensemble can be adjusted to allow for time-sharing as follows. Fix a finite time-sharing alphabet $\mathcal{U}$, a time-sharing sequence $\boldsymbol{u} \in \mathcal{U}^n$ with empirical distribution Q_U, and the input distributions $Q_1(x_1|u)$ and $Q_2(x_2|u)$. The cost constraints are of the same form as (53), where the empirical average of $a(u,x_\nu)$ is constrained to be close to its average $\mathbb{E}[a(U,X_\nu)]$. Following the proof of Theorem 3, we obtain E_0 functions of the form $E_0=\sum_u Q_U(u)E_0^{(u)}$, where $E_0^{(u)}$ is the E_0 function in Theorem 3 associated with $Q_1(\cdot|u)$ and $Q_2(\cdot|u)$.

Setting a_1 and a_2 to zero in each of (58)–(60) recovers Gallager's error exponent for i.i.d. random coding [18]. The following lemma shows that, after the optimization of a_1 and a_2, our exponent is equivalent to that of Liu and Hughes [6], which was obtained using constant-composition random coding. We present the lemma in the absence of time-sharing, but the same statement holds when we consider the more general result outlined in Remark 1.

Lemma 1. *The exponents given in* (55)–(57) *coincide with* [6, Theorem 1] *under* $|\mathcal{U}|=1$.

Proof: Since the proof is rather detailed but based on existing techniques, we provide only an outline. Each of the three exponents are handled similarly, so we focus on $E_{r,12}$. We denote the corresponding exponent in [6] by $\hat{E}_{r,12}(\boldsymbol{Q},R_1,R_2)$. From [6, Eq. (30)], we can write $\hat{E}_{r,12}$ in the form $\max_{\rho\in[0,1]} \hat{E}_{0,12}(\boldsymbol{Q},\rho)-\rho(R_1+R_2)$, where

$$\hat{E}_{0,12}(\boldsymbol{Q},\rho) \triangleq \min_{\substack{\widetilde{P}_{X_1X_2Y}:\\ \widetilde{P}_{X_1}=Q_1,\widetilde{P}_{X_2}=Q_2}} D\big(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times W\big)+\rho D\big(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times \widetilde{P}_Y\big). \quad (64)$$

We split the minimization by first minimizing over $\widetilde{P}_{X_1X_2Y}$ subject to $\widetilde{P}_{X_1Y}=\hat{P}_{X_1Y}$ and $\widetilde{P}_{X_2}=Q_2$, and then minimizing over $\hat{P}_{X_1Y}$ subject to $\hat{P}_{X_1}=Q_1$. Starting with the former minimization and using the identity

$$D\big(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times \widetilde{P}_Y\big)=I_{\widetilde{P}}(X_1;Y)+I_{\widetilde{P}}(X_2;X_1,Y), \quad (65)$$

it can be shown that

$$\hat{E}_{0,12}(\boldsymbol{Q},\rho)=\min_{\hat{P}_{X_1Y}:\hat{P}_{X_1}=Q_1}\sup_{a_2'(\cdot)} H_{\hat{P}}(Y|X_1) + \rho I_{\hat{P}}(X_1;Y)-(1+\rho)\sum_{x_1,y}\hat{P}_{X_1Y}(x_1,y) \times \log\sum_{x_2} Q_2(x_2)W(y|x_1,x_2)^{\frac{1}{1+\rho}} e^{\frac{a_2'(x_2)-\phi_2'}{1+\rho}}, \quad (66)$$

where the supremum over a_2' is over all real-valued functions on $\mathcal{X}_2$, and $\phi_2' \triangleq \mathbb{E}_Q[a_2'(X_2)]$. Specifically, one can upper bound $\hat{E}_{0,12}$ by the right-hand side of (64) using Lagrange duality, and a matching lower bound can be obtained using the log-sum inequality, similarly to [20, Appendix A].

Using Fan's minimax theorem [21], we can swap the order of the minimum and the supremum in (66). Writing the mutual information $I_{\hat{P}}(X;Y)$ in the form [8]

$$\min_{Q_Y}\sum_{x,y}\hat{P}_{X_1Y}(x_1,y)\log\frac{\hat{P}_{X_1Y}(x_1,y)}{Q_1(x_1)Q_Y(y)} \tag{67}$$

we can again using Lagrange duality techniques to show that

$$\hat{E}_{0,12}(\boldsymbol{Q},\rho) = \sup_{a_2'(\cdot)}\min_{Q_Y} -(1+\rho)\sum_{x_1}Q_1(x_1)$$
$$\times\log\sum_{x_2,y}Q_2(x_2)W(y|x_1,x_2)^{\frac{1}{1+\rho}}e^{\frac{a_2'(x_2)-\phi_2'}{1+\rho}}Q_Y(y). \tag{68}$$

We can write the objective in (68) as

$$-(1+\rho)D(Q_1\|\widetilde{Q}_1)-(1+\rho)\sum_{x_1}Q_1(x_1)$$
$$\times\log\frac{\widetilde{Q}_1(x_1)}{Q_1(x_1)}\sum_{x_2,y}Q_2(x_2)W(y|x_1,x_2)^{\frac{1}{1+\rho}}e^{\frac{a_2'(x_2)-\phi_2'}{1+\rho}}Q_Y(y), \tag{69}$$

where $\widetilde{Q}_1$ is an arbitrary distribution with the same support as Q. Using this expression and following similar steps to [5, Sec. 2.5, Prob. 23], we obtain

$$\hat{E}_{0,12}(\boldsymbol{Q},\rho) = \sup_{a_2'(\cdot)}\max_{\widetilde{Q}_1} -(1+\rho)D(Q_1\|\widetilde{Q}_1)$$
$$-\log\sum_{y}\left(\sum_{x_1}\widetilde{Q}_1(x_1)Q_2(x_2)W(y|x_1,x_2)^{\frac{1}{1+\rho}}e^{\frac{a_2'(x_2)-\phi_2'}{1+\rho}}\right)^{1+\rho}, \tag{70}$$

where $\widetilde{Q}_1$ is constrained to have the same support as Q. Finally, using the techniques of [15, Theorems 4-5], it can be shown that (70) yields

$$\hat{E}_{0,12}(\boldsymbol{Q},\rho) = \sup_{a_1(\cdot),a_2'(\cdot)} -\log\sum_{y}\left(\sum_{x_1,x_2}Q_1(x_1)Q_2(x_2)\right.$$
$$\left.\times W(y|x_1,x_2)^{\frac{1}{1+\rho}}e^{a_1(x_1)-\phi_1}e^{\frac{a_2'(x_2)-\phi_2'}{1+\rho}}\right)^{1+\rho}. \tag{71}$$

The proof is concluded by identifying $a_2(x_2)=\frac{a_2'(x_2)}{1+\rho}$. ■

In contrast to the single-user setting, the error exponents for constant-composition random coding [6] can be strictly greater than that of i.i.d. random coding [18] even after the optimization of the input distributions; see [6] for details. Thus, the exponent of Theorem 3 can be strictly higher than that of [18]. A further advantage of the exponent in Theorem 3 is that it can be applied to channels with infinite or continuous alphabets. In contrast, the analysis of [6] is based on the method of types, and is valid only for finite alphabets.

REFERENCES

[1] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, July and Oct. 1948.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2001.

[3] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[4] R. Fano, *Transmission of information: A statistical theory of communications*. M.I.T. Press, 1961.

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

[6] Y. Liu and B. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, March 1996.

[7] 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.

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

[9] 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.

[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] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, 2008, vol. 2.

[12] T. S. Han, *Information-spectrum methods in information theory*. Springer, 2002.

[13] T. Tao, *Topics in Random Matrix Theory*. American Mathematical Society, 2012.

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

[15] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Ensemble-tight error exponents for mismatched decoders," in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.

[16] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," http://arxiv.org/abs/1109.6310.

[17] J. Scarlett and A. Guillén i Fàbregas, "An achievable error exponent for the mismatched multiple-access channel," in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.

[18] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, March 1985.

[19] 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.

[20] 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.

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