A Dual-Domain Achievability of the Typical Error Exponent

Giuseppe Cocco
Universitat Pompeu Fabra
giuseppe.cocco@upf.edu

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

Josep Font-Segura
Universitat Pompeu Fabra
josep.font@ieee.org

***Abstract*—For random-coding ensembles with pairwise-independent codewords, we show that the probability that the exponent of a given code from the ensemble being smaller than an upper bound on the typical random-coding exponent is vanishingly small. This upper bound is known to be tight for i.i.d. ensembles over the binary symmetric channel and for constant-composition codes over memoryless channels. Our result recovers these as special cases and remains valid for arbitrary alphabets and channel memory, as well as arbitrary ensembles with pairwise independent codewords.**

I. Introduction

We consider coding over discrete channels with conditional probability distribution $W^n(\boldsymbol{y}|\boldsymbol{x})$, being $\boldsymbol{x} \in \mathcal{X}^n, \boldsymbol{y} \in \mathcal{Y}^n$ the transmitted and received sequences of length n, and $\mathcal{X}, \mathcal{Y}$ the finite channel input and output alphabets, respectively. For memoryless channels we have $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W(y_i|x_i)$, where $x_i \in \mathcal{X}, y_i \in \mathcal{Y}$. A code $\mathcal{C}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M_n}\}$ is a set of M_n codewords of length n. Assuming equiprobable messages and maximum-likelihood decoding, the error probability of a fixed code $\mathcal{C}_n$ is given by $P_{\mathrm{e}}(\mathcal{C}_n) \triangleq \frac{1}{M_n}\sum_{m=1}^{M_n} P_{\mathrm{e}}(\mathcal{C}_n, m)$, where $P_{\mathrm{e}}(\mathcal{C}_n, m)$ is the error probability when codeword $\boldsymbol{x}_m$ is transmitted. We define the error exponent of a code $\mathcal{C}_n$ as

$$E_n(\mathcal{C}_n) \triangleq -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n). \tag{1}$$

An exponent E is achievable when there exists a sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ such that $\liminf_{n\to\infty} E_n(\mathcal{C}_n) \geq E$.

Lower bounds on the error exponent of codes used over discrete memoryless channels are traditionally derived using random-coding arguments [1, Sec. 5.6], [2, Ch. 10]. Let C_n be the random variable representing a code randomly generated with some probability distribution. In code ensembles with pairwise-independent codewords, such as the i.i.d. or the constant composition ensembles later discussed in Sec. III, M_n codewords are generated independently with probability distribution $Q^n(\boldsymbol{x})$. For such ensembles and sufficiently large n, there exists a sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ whose error exponent (1) is at least as large as the random-coding error exponent $E_{\mathrm{r}}(R, Q)$ given by

$$E_{\mathrm{r}}(R, Q) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}[P_{\mathrm{e}}(\mathsf{C}_n)], \tag{2}$$

where $R \triangleq \lim_{n\to\infty}\frac{1}{n}\log M_n$ is the code rate and Q is the asymptotic single-letter version of Q^n. It is known that (2) is smaller than the reliability function for low code rates and was improved by Gallager by employing expurgation [1, Sec. 5.7]. Expurgation shows the existence of a codebook with an improved exponent, the expurgated exponent $E_{\mathrm{ex}}(R, Q)$ [3, Eq. (5.7.11)].

In contrast, Barg and Forney [4] and Merhav [5], considered the typical random-coding error exponent, defined it as the limiting expected error exponent of the ensemble

$$E_{\mathrm{trc}}(R, Q) \triangleq \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\mathrm{e}}(\mathsf{C}_n)]. \tag{3}$$

The typical error exponent improves over the random-coding error exponent (2) at low rates, and is achieved by codes in the specified ensemble, unlike the expurgated exponent, achieved by codes with unknown structure. For the i.i.d. [4] and the constant-composition [5] [6, Lemma 3] ensembles, the typical error exponent is shown to satisfy $E_{\mathrm{trc}}(R, Q) = E_{\mathrm{ex}}(2R, Q) + R \leq E_{\mathrm{ex}}(R, Q)$, with equality for $R = 0$ [1]. In the constant-composition case, it is further known that the probability of finding codes outperforming the typical error exponent is double-exponentially decaying in n [8].

II. Main Result

The main result of this paper is a consequence of the lemma below, a refinement of a Lemma by Gallager [1, p. 151] that can be shown by applying Markov's inequality to the random variable $P_{\mathrm{e}}(\mathsf{C}_n)^s$ for $s > 0$.

Lemma 1. *Let γ_n be a real-valued monotonically increasing sequence in n such that $\lim_{n\to\infty}\gamma_n = \infty$ and $\lim_{n\to\infty}\frac{1}{n}\log\gamma_n = 0$. For an arbitrary random-coding ensemble and $s > 0$, it holds that*

$$\mathbb{P}\Big[P_{\mathrm{e}}(\mathsf{C}_n) \geq \gamma_n^{\frac{1}{s}}\mathbb{E}[P_{\mathrm{e}}(\mathsf{C}_n)^s]^{\frac{1}{s}}\Big] \leq \frac{1}{\gamma_n}. \tag{4}$$

This work has been funded in part by the postdoctoral fellowship programme Beatriu de Pinós, funded by the Secretary of Universities and Research (Government of Catalonia), by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 801370, and by the European Research Council under ERC grant agreement 725411.

[1] In [5] an inequality sign is used, but this is only because the improved expurgated presented in [7, Section 1, point 4.] is used instead of Gallager's.

Lemma 1 implies that the probability to randomly generate a code $\mathcal{C}_n$ such that

$$P_{\mathrm{e}}(\mathcal{C}_n) < \gamma_n^{\frac{1}{s}} \mathbb{E}[P_{\mathrm{e}}(\mathsf{C}_n)^s]^{\frac{1}{s}} \tag{5}$$

is larger than $1 - \frac{1}{\gamma_n}$. The r.h.s. of (5) shows a strong connection with the typical random-coding exponent (3) as $\lim_{s\to 0} \mathbb{E}[P_{\mathrm{e}}(\mathsf{C}_n)^s]^{\frac{1}{s}} = \mathbb{E}[\log P_{\mathrm{e}}(\mathsf{C}_n)]$. However, since the lemma assumes that $\gamma_n > 1$ from a certain n, the bound (5) is tightened for $s < \infty$, rather leading to a lower bound on the error exponent of a typical random code. This result is subsumed in the following theorem, valid for channels with arbitrary alphabets or memory and pairwise-independent code ensembles, thus significantly extending the results in [4]–[6].

Theorem 1. *For a channel W^n and a pairwise-independent ensemble with codeword distribution Q^n, it holds that*

$$\lim_{n\to\infty} \mathbb{P}\big[E_n(\mathsf{C}_n) \geq E_{\mathrm{ex}}^n(2R, Q^n) + R - \delta_n\big] = 1, \tag{6}$$

where $\delta_n > 0$ such that $\lim_{n\to\infty} \delta_n = 0$,

$$E_{\mathrm{ex}}^n(R, Q^n) = E_{\mathrm{x}}^n(\hat{\rho}_n, Q^n) - \hat{\rho}_n R \tag{7}$$

is the multi-letter version of the expurgated exponent,

$$E_{\mathrm{x}}^n(\rho, Q^n) \triangleq -\frac{1}{n} \log\Bigg(\sum_{\boldsymbol{x}} \sum_{\boldsymbol{x}'} Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}') Z_n(\boldsymbol{x}, \boldsymbol{x}')^{\frac{1}{\rho}} \Bigg)^{\rho}, \tag{8}$$

$Z_n(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\boldsymbol{y}} \sqrt{W^n(\boldsymbol{y}|\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}')}$ is the Bhattacharyya coefficient between $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$ and

$$\hat{\rho}_n = \arg\max_{\rho \geq 1} \big\{ E_{\mathrm{x}}^n(\rho, Q^n) - \rho 2R \big\} \tag{9}$$

is the bound parameter that yields the highest exponent.

Proof. We start deriving an upper bound on the average tilted error probability. For a given code $\mathcal{C}_n$ we have:

$$P_{\mathrm{e}}(\mathcal{C}_n) = \frac{1}{M_n} \sum_{m=1}^{M_n} P_{\mathrm{e}}(\mathcal{C}_n, m) \tag{10}$$

$$\leq \frac{1}{M_n} \sum_{m=1}^{M_n} \sum_{m' \neq m} Z_n(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \tag{11}$$

where (11) follows from the Bhattacharyya bound [9, Sec. 2.3] expressed in terms of the Bhattacharyya coefficient $Z_n(\boldsymbol{x}, \boldsymbol{x}')$. For $s \in (0, 1]$, we bound $P_{\mathrm{e}}(\mathcal{C}_n)^s$ as

$$P_{\mathrm{e}}(\mathcal{C}_n)^s \leq \left(\frac{1}{M_n} \sum_{m=1}^{M_n} \sum_{m' \neq m} Z_n(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \right)^s \tag{12}$$

$$\leq \sum_{m=1}^{M_n} \sum_{m' \neq m} \left(\frac{1}{M_n} Z_n(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \right)^s. \tag{13}$$

Taking the ensemble average at both sides in (13), we obtain

$$\mathbb{E}[P_{\mathrm{e}}(\mathsf{C}_n)^s] \leq \frac{1}{M_n^s} \sum_{m=1}^{M_n} \sum_{m' \neq m} \mathbb{E}\Big[Z_n(\boldsymbol{X}_m, \boldsymbol{X}_{m'})^s \Big], \tag{14}$$

where $\boldsymbol{X}_m, \boldsymbol{X}_{m'}$ are randomly generated codewords corresponding to messages m, m'. For pairwise-independent ensembles, the expectation over the ensemble does not depend on the specific pair of codewords, and (14) simplifies to

$$\mathbb{E}[P_{\mathrm{e}}(\mathsf{C}_n)^s] \leq \frac{1}{M_n^s} M_n(M_n - 1) \mathbb{E}\Big[Z_n(\boldsymbol{X}, \boldsymbol{X}')^s \Big], \tag{15}$$

where $\boldsymbol{X}$ and $\boldsymbol{X}'$ have joint distribution $Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}')$.

Using the r.h.s. of (15) in (5) and making the change of variable $s = \frac{1}{\rho}$ for $\rho \geq 1$, we obtain

$$P_{\mathrm{e}}(\mathcal{C}_n) \leq \frac{1}{M_n} \big(\gamma_n M_n(M_n - 1) \big)^{\rho} \left(\mathbb{E}\Big[Z_n(\boldsymbol{X}, \boldsymbol{X}')^{\frac{1}{\rho}} \Big] \right)^{\rho}. \tag{16}$$

Taking the negative normalized logarithm of both sides in (16), we lower bound the error exponent of a code $\mathcal{C}_n$ (1) as

$$E_n(\mathcal{C}_n) \geq E_{\mathrm{x}}^n(\rho, Q^n) - \rho 2R + R - \frac{\rho}{n} \log \gamma_n, \tag{17}$$

where we defined

$$E_{\mathrm{x}}^n(\rho, Q^n) \triangleq -\frac{1}{n} \log\Bigg(\sum_{\boldsymbol{x}} \sum_{\boldsymbol{x}'} Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}') Z_n(\boldsymbol{x}, \boldsymbol{x}')^{\frac{1}{\rho}} \Bigg)^{\rho} \tag{18}$$

as the multiletter version of the Gallager's E_{x}-function [3, Eq. (5.7.12)], and used that $\log M_n = nR$. For every n, we let $\hat{\rho}_n$ be the optimizing parameter defined in (9). Then, the lower bound (17) is further tightened as

$$E_n(\mathcal{C}_n) \geq E_{\mathrm{ex}}^n(2R, Q^n) + R - \delta_n, \tag{19}$$

where $E_{\mathrm{ex}}^n(R, Q^n)$ is the multi-letter version of the expurgated exponent [3, Eq. (5.7.11)] given by

$$E_{\mathrm{ex}}^n(R, Q^n) = E_{\mathrm{x}}^n(\hat{\rho}_n, Q^n) - \hat{\rho}_n R \tag{20}$$

and

$$\delta_n = \frac{\hat{\rho}_n}{n} \log \gamma_n. \tag{21}$$

Using Lemma 1 for the ensembles of codes whose error probability satisfy (16), we obtain that

$$\mathbb{P}\big[E_n(\mathsf{C}_n) \geq E_{\mathrm{ex}}^n(2R, Q^n) + R - \delta_n\big] > 1 - \frac{1}{\gamma_n}. \tag{22}$$

Define $\hat{\rho} = \lim_{n\to\infty} \hat{\rho}_n$. We have that $\lim_{n\to\infty} \delta_n = \lim_{n\to\infty} \frac{1}{n} \log \gamma_n = 0$ by definition of γ_n when $\hat{\rho} < \infty$. Instead, for $\hat{\rho}_n \to \infty$ we have that $\lim_{n\to\infty} \delta_n = 0$ if and only if $\hat{\rho}_n$ grows slower than $\frac{n}{\log \gamma_n}$, e.g., as $\hat{\rho}_n = \frac{\sqrt{n}}{\log \gamma_n}$. Under these assumptions, δ_n vanishes with n. Hence, taking the limit $n \to \infty$ at both sides of (22) we obtain (6). □

For all channels W^n and ensembles with pairwise-independent codeword distribution Q^n such that the limit in (6) can be brought inside the square brackets, Theorem 1 reads

$$\mathbb{P}\Big[\lim_{n\to\infty} \{E_n(\mathsf{C}_n) + \delta_n\} \geq E_{\mathrm{ex}}(2R, Q) + R \Big] = 1, \tag{23}$$

where $E_{\mathrm{ex}}(R, Q)$ is the asymptotic single-letter version of the expurgated exponent (20). This is guaranteed if (i) the Borel-Cantelli lemma [10] applies and (ii) we have that $\lim_{n\to\infty} E_{\mathrm{ex}}^n(2R, Q^n) = E_{\mathrm{ex}}(2R, Q)$, which is the case, for instance, in memoryless channels. The result in (23) implies

that the achievable typical random-coding exponent is lower bounded as $E_{\rm ex}(2R,Q)+R$. Such lower bound is known to be tight for constant-composition ensembles [5] and coincides with the one derived in [4] for the binary symmetric channel (BSC).

Observe that Theorem 1 does not apply to code ensembles whose codeword generation is not pairwise independent, such as the random Gilbert-Varshamov (RGV) ensemble [11]. For this ensemble, it is known that the exponent of the ensemble average error probability is the maximum of the expurgated and random-coding exponents [11]. For ensembles whose average error probability achieves the expurgated exponent, like the RGV, by setting $s=1$ in Lemma 1, we find that

$$\mathbb{P}\Big[E_n(\mathsf{C}_n)\geq E_{\rm ex}(R,Q)-\delta_n\Big]>1-\epsilon_n, \tag{24}$$

implying that the error exponent of typical codes is lower bounded by $E_{\rm ex}(R,Q)$. This can also be seen as a consequence of Jensen's inequality, since $\mathbb{E}[\log P_{\rm e}(\mathsf{C}_n)]\leq \log\mathbb{E}[P_{\rm e}(\mathsf{C}_n)] = E_{\rm ex}(R,Q)$, and thus $E_{\rm trc}(R,Q) = \lim_{n\to\infty}-\frac{1}{n}\mathbb{E}[\log P_{\rm e}(\mathsf{C}_n)]\geq E_{\rm ex}(R,Q)$.

III. I.I.D. and Constant Composition Ensembles

We now specialize Theorem 1 to the i.i.d. and constant-composition ensembles over memoryless channels under a common single-letter input distribution $Q(x)$. As a by-product, we provide a direct dual-domain derivation of the expurgated exponent for constant composition-codes [12], as an alternative to that found in [13].

A. i.i.d. Ensemble

We first consider the i.i.d. ensemble with distribution

$$Q_{\rm iid}^n(\boldsymbol{x})=\prod_{i=1}^n Q(x_i). \tag{25}$$

Using (25) in (18), we recover the single-letter version of the Gallager's $E_{\rm x}$-function [3, Eq. (5.7.12)], namely

$$E_{\rm x}^{\rm iid}(\rho,Q)=-\rho\log\sum_x\sum_{x'}Q(x)Q(x')Z(x,x')^{\frac{1}{\rho}}, \tag{26}$$

where $Z(x,x')$ is the single-letter Bhattacharyya coefficient, and the i.i.d. expurgated exponent in (20) becomes

$$E_{\rm ex}^{\rm iid}(R,Q)=E_{\rm x}^{\rm iid}(\hat\rho,Q)-\hat\rho R. \tag{27}$$

When $R>0$, it follows that the optimal parameter (9) does not depend on n and reads

$$\hat\rho=\arg\max_{\rho\geq 1}\big\{E_{\rm x}^{\rm iid}(\rho,Q)-\rho 2R\big\}. \tag{28}$$

In this case, since $\hat\rho<\infty$, the condition $\lim_{n\to\infty}\delta_n=0$ of the theorem is satisfied. For strictly zero rates, namely when $\lim_{n\to\infty}\frac{1}{n}\log M_n=0$, the largest exponent in (27) is achieved for $\hat\rho\to\infty$. We then consider the dependence on n in the rate setting $R=\frac{1}{n}\log M_n$ in (28) and let $\hat\rho_n$ be

$$\hat\rho_n=\arg\max_{\rho\geq 1}\Big\{E_{\rm x}^{\rm iid}(\hat\rho,Q)-\frac{2\rho}{n}\log M_n\Big\}. \tag{29}$$

In order to characterize how $\hat\rho_n$ grows as $n\to\infty$, we find the Taylor series expansion of $E_{\rm x}^{\rm iid}(\rho,Q)$ around $\rho\to\infty$, yielding

$$E_{\rm x}^{\rm iid}(\rho,Q)=a_0(Q)-\frac{a_1(Q)}{\rho}+O(\rho^{-2}), \tag{30}$$

where

$$a_0(Q)=-\sum_x\sum_{x'}Q(x)Q(x')\log Z(x,x') \tag{31}$$

and

$$a_1(Q)=\frac{1}{2}\left(\sum_x\sum_{x'}Q(x)Q(x')\log^2 Z(x,x')-a_0^2(Q,W)\right). \tag{32}$$

Then, it follows that $\hat\rho_n$ grows as

$$\hat{\rho_n}=\sqrt{\frac{n\cdot a_1(Q)}{2\log M_n+\log\gamma_n}}+\sigma_n, \tag{33}$$

where σ_n is a term that vanishes as $n\to\infty$, implying that δ_n defined in (21) satisfies $\lim_{n\to\infty}\delta_n=0$ as in the $R>0$ case.

According to Theorem 1, with probability approaching one as $n\to\infty$, an i.i.d. code $\mathcal{C}_n^{\rm iid}$ randomly generated from the ensemble (25) has an error exponent (1) satisfying

$$E_n(\mathcal{C}_n^{\rm iid})\geq E_{\rm ex}^{\rm iid}(2R,Q)+R-\delta_n \tag{34}$$

with vanishing δ_n. Particularizing (27) to the BSC case, it follows that the bound (34) coincides with the TRC exponent given by Barg and Forney in [4].

B. Constant-Composition Ensemble

For every n, let $\hat Q_n$ be a type, or empirical distribution, such that $\|\hat Q_n-Q\|_\infty\leq\frac{1}{n}$ where $\|P\|_\infty=\max_x P(x)$. Then, the constant-composition ensemble has codeword distribution

$$Q_{\rm cc}^n(\boldsymbol{x})=\frac{1}{|\mathcal{T}^n(\hat Q_n)|}\mathbb{1}\big\{\boldsymbol{x}\in\mathcal{T}^n(\hat Q_n)\big\}, \tag{35}$$

where $\mathcal{T}_n(\hat Q_n)$ is the type class, i.e., the set of all sequences of length n with empirical distribution $\hat Q_n$, while $\mathbb{1}\{.\}$ is the indicator function.

For the constant-composition ensemble (35), the normalized multi-letter Gallager's expurgated function $E_{\rm x}^n(\rho,Q^n)$ reads

$$E_{\rm x}^n(\rho,Q^n)=-\frac{1}{n}\log\left(\sum_{\boldsymbol{x}}Q^n(\boldsymbol{x})\sum_{\boldsymbol{x}'}Q^n(\boldsymbol{x}')\prod_{i=1}^n Z(x_i,x_i')^{\frac{1}{\rho}}\right)^\rho. \tag{36}$$

In (36), we have used the fact that the channel is memoryless to express $Z_n(\boldsymbol{x},\boldsymbol{x}')$ as a product, and we have also changed the order of the summations over $\boldsymbol{x}$ and $\boldsymbol{x}'$.

Since all codewords $\boldsymbol{x}'$ have the same probability $\frac{1}{|\mathcal{T}^n(\hat Q_n)|}$, the summation over $\boldsymbol{x}'$ in (36) for a fixed $\boldsymbol{x}$ satisfies

$$\sum_{\boldsymbol{x}'}Q^n(\boldsymbol{x}')\prod_{i=1}^n Z(x_i,x_i')^{\frac{1}{\rho}}=\frac{1}{|\mathcal{T}^n(\hat Q_n)|}\sum_{\boldsymbol{x}'}\prod_{i=1}^n Z(x_i,x_i')^{\frac{1}{\rho}}. \tag{37}$$

Identifying $g_i(x_i)=Z(x_i,x_i')^{\frac{1}{\rho}}$ in [14, Eq. (2.4)], the r.h.s. of (37) can be further upper bounded by

$$\min_{\bar P}\left\{(n+1)^{|\mathcal{X}|-1}e^{nD(\hat Q_n\|\bar P)}\prod_{i=1}^n\sum_{x'}\bar P(x')Z(x_i,x')^{\frac{1}{\rho}}\right\}, \tag{38}$$

where $\bar{P}$ is an auxiliary probability distribution and $D(Q\|\bar{P})$ is the relative entropy between distributions Q and $\bar{P}$. Using the upper bound (38) into (36) and arranging terms, we obtain

$$E_{\mathrm{x}}^{n}(\rho, Q^n) \geq \max_{\bar{P}} \Bigg\{ -\frac{1}{n} \log \Bigg((n+1)^{|\mathcal{X}|-1} e^{nD(\hat{Q}_n\|\bar{P})} \cdot \cdot \sum_{\boldsymbol{x}} Q^n(\boldsymbol{x}) \prod_{i=1}^{n} \sum_{x'} \bar{P}(x') Z(x_i, x')^{\frac{1}{\rho}} \Bigg)^{\rho} \Bigg\}. \quad (39)$$

Since, again, all codewords $\boldsymbol{x}$ have the same probability, (39) simplifies to

$$E_{\mathrm{x}}^{n}(\rho, Q^n) \geq \max_{\bar{P}} \Bigg\{ -\frac{1}{n} \log \Bigg((n+1)^{|\mathcal{X}|-1} e^{nD(\hat{Q}_n\|\bar{P})} \cdot \cdot \frac{1}{|\mathcal{T}^n(\hat{Q}_n)|} \sum_{\boldsymbol{x}} \prod_{i=1}^{n} \sum_{x'} \bar{P}(x') Z(x_i, x')^{\frac{1}{\rho}} \Bigg)^{\rho} \Bigg\}. \quad (40)$$

The sum over $\boldsymbol{x}$ in (40) has exactly $|\mathcal{T}^n(\hat{Q}_n)|$ terms, hence,

$$E_{\mathrm{x}}^{n}(\rho, Q^n) \geq \max_{\bar{P}} \Bigg\{ -\frac{1}{n} \log \Bigg((n+1)^{|\mathcal{X}|-1} e^{nD(\hat{Q}_n\|\bar{P})} \cdot \cdot \prod_{x} \Bigg(\sum_{x'} \bar{P}(x') Z(x, x')^{\frac{1}{\rho}} \Bigg)^{n\hat{Q}_n(x)} \Bigg)^{\rho} \Bigg\}, \quad (41)$$

where we used elementary properties of constant-composition sequences. Next, we rewrite the maximization over $\bar{P}$ as maximization over an auxiliary function $a(x)$ [13] satisfying $\bar{P}(x) = Q(x) e^{\frac{a(x)}{\rho}}$. Using this in (41), we find

$$E_{\mathrm{x}}^{n}(\rho, Q^n) \geq \max_{a(x)} \Bigg\{ -\frac{1}{n} \log \Bigg((n+1)^{|\mathcal{X}|-1} e^{-\frac{n}{\rho} \sum_x \hat{Q}_n(x) a(x)} \cdot \cdot \prod_{x} \Bigg(\sum_{x'} Q(x') e^{\frac{a(x')}{\rho}} Z(x, x')^{\frac{1}{\rho}} \Bigg)^{n\hat{Q}_n(x)} \Bigg)^{\rho} \Bigg\}, \quad (42)$$

which, after some algebra, can be further simplified to

$$E_{\mathrm{x}}^{n}(\rho, Q^n) \geq -\frac{\rho(|\mathcal{X}|+1)}{n} \log(n+1) + \max_{a(x)} \Bigg\{ -\rho \log \prod_{x} \Bigg(\sum_{x'} \hat{Q}_n(x') \Big(Z(x, x') e^{a(x') - a(x)} \Big)^{\frac{1}{\rho}} \Bigg)^{\hat{Q}_n(x)} \Bigg\}. \quad (43)$$

We note that the the first term in the r.h.s. of (43) vanishes with n. Using that $\|\hat{Q}_n - Q\|_\infty \leq \frac{1}{n}$, and further simplifying the result, we obtain the constant-composition version of the Gallager's E_{x}-function, namely

$$E_{\mathrm{x}}^{\mathrm{cc}}(\rho, Q) = \max_{a(x)} \Bigg\{ -\rho \sum_{x} Q(x) \cdot \cdot \log \sum_{x'} Q(x') \Big(Z(x, x') e^{a(x') - a(x)} \Big)^{\frac{1}{\rho}} \Bigg\}. \quad (44)$$

According to Theorem 1, with probability approaching one as $n \to \infty$, a constant-composition code $\mathcal{C}_n^{\mathrm{cc}}$ randomly generated from the ensemble (35) has an error exponent (1) satisfying the typical random-coding lower bound [5]

$$E_n(\mathcal{C}_n^{\mathrm{cc}}) \geq E_{\mathrm{ex}}^{\mathrm{cc}}(2R, Q) + R, \quad (45)$$

where

$$E_{\mathrm{ex}}^{\mathrm{cc}}(R, Q) = E_{\mathrm{x}}^{\mathrm{cc}}(\hat{\rho}, Q) - \hat{\rho} R \quad (46)$$

is the constant-composition version of the expurgated exponent with optimal parameter

$$\hat{\rho} = \arg\max_{\rho \geq 1} \big\{ E_{\mathrm{x}}^{\mathrm{cc}}(\rho, Q) - \rho 2R \big\}. \quad (47)$$

IV. Conclusion

We studied the typical random-coding error exponent by means of a refinement of a lemma by Gallager. We showed that the probability that the exponent of a code from a pairwise-independent ensemble is smaller than $E_{\mathrm{ex}}^{n}(2R, Q^n) + R$ vanishes with n. The method provides a new, dual-domain achievability of the typical random-coding exponent and explicitly establishes a connection between typical random coding and the expurgated method as pointed out in [4].

Theorem 1 is valid for pairwise-independent ensembles and channels with arbitrary alphabets and memory, significantly extending the results in the literature that are limited to binary codes over the BSC [4] and constant-composition codes over memoryless channels [5]. For memoryless channels, our results in Section III hold for discrete or continuous inputs and outputs, with the exception of the constant-composition ensemble that requires a discrete input.

References

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

[2] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

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

[4] A. Barg and G. D. Forney, "Random codes: minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, Sep. 2002.

[5] N. Merhav, "Error exponents of typical random codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6223–6235, Sep. 2018.

[6] A. Nazari, A. Anastasopoulos, and S. Pradhan, "Error exponent for multiple-access channels: Lower bounds," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5095–5115, 2014.

[7] N. Merhav, "The generalized stochastic likelihood decoder: Random coding and expurgated bounds," *IEEE Trans. on Info. Theory*, vol. 63, no. 8, pp. 5039–5051, Aug. 2017.

[8] R. Tamir, N. Merhav, N. Weinberger, and A. Guillén i Fàbregas, "Large deviations behavior of the logarithmic error probability of random codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6635–6659, 2020.

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

[10] R. Durrett, *Probability: Theory and Examples*, 5th ed. Cambridge University Press, Apr. 2019.

[11] A. Somekh-Baruch, J. Scarlett, and A. Guillén i Fàbregas, "Generalized random Gilbert-Varshamov codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3452–3469, 2019.

[12] 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.

[13] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements, and connections," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4449–4462, 2014.

[14] G. Poltyrev, "Random coding bounds for discrete memoryless channels," *Probl. Peredachi Inf.*, vol. 18, no. 1, pp. 12–26, 1982.