A Derivation of the Asymptotic Random-Coding Prefactor

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 the subexponential prefactor to the random-coding bound for a given rate. Using a refinement of Gallager's bounding techniques, an alternative proof of a recent result by Altuğ and Wagner is given, and the result is extended to the setting of mismatched decoding.**

I. Introduction

Error exponents are a widely-studied tool in information theory for characterizing the performance of coded communication systems. Early works on error exponents for discrete memoryless channels (DMCs) include those of Fano [1, Ch. 9], Gallager [2, Ch. 5] and Shannon *et al.* [3]. The achievable exponent of [1], [2] was obtained using i.i.d. random coding, and coincides with the sphere-packing exponent given in [3] for rates above a threshold called the critical rate.

Denoting the exponent of [1], [2] by $E_r(R)$, we have the following: For all (n,R), there exists a code of rate R and block length n such that the error probability p_e satisfies $p_e \le \alpha(n,R)e^{-nE_r(R)}$, where $\alpha(n,R)$ is a subexponential prefactor. In both [1] and [2], the prefactor is $O(1)$. In particular, Gallager showed that one can achieve $\alpha(n,R)=1$.

Early works on improving the $O(1)$ prefactor for certain channels and rates include those of Elias [4], Dobrushin [5] and Gallager [6]. These results were recently generalized by Altuğ and Wagner [7]–[9], who obtained prefactors to the random-coding bound at all rates below capacity, as well as converse results above the critical rate. The bounds in [7], [8] were obtained using i.i.d. random coding, and the behavior of the prefactor varies depending on whether the rate is above or below the critical rate, and whether a regularity condition is satisfied (see Section II).

In this paper, we give an alternative proof of the main result of [7], [8], as well as a generalization to the setting of mismatched decoding [10]–[14], where the decoding rule is fixed and possibly suboptimal (e.g. due to channel uncertainty or implementation constraints). The analysis of [7], [8] can be considered a refinement of that of Fano [1, Ch. 9], whereas the analysis in this paper can be considered a refinement of that of Gallager [2, Ch. 5]. Our techniques can also be used to derive Gallager's expurgated exponent [2, Ch. 5.7] with an $O\big(\frac{1}{\sqrt{n}}\big)$ prefactor under some technical conditions [15], thus improving on Gallager's $O(1)$ prefactor.

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.

A. Notation

Vectors are written using bold symbols (e.g. $\boldsymbol{x}$), and the corresponding i-th entry is written with a subscript (e.g. x_i). For two sequences f_n and g_n, we write $f_n = O(g_n)$ if $|f_n| \le c|g_n|$ for some c and sufficiently large n, and $f_n = o(g_n)$ if $\lim_{n\to\infty}\frac{f_n}{g_n}=0$. The indicator function is denoted by $\mathbb{1}\{\cdot\}$.

The marginals of a joint distribution $P_{XY}(x,y)$ are denoted by $P_X(x)$ and $P_Y(y)$. Expectation with respect to a joint distribution $P_{XY}(x,y)$ is denoted by $\mathbb{E}_P[\cdot]$, or simply $\mathbb{E}[\cdot]$ when the probability distribution is understood from the context. Given a distribution $Q(x)$ and conditional distribution $W(y|x)$, we write $Q\times W$ to denote the joint distribution defined by $Q(x)W(y|x)$. The set of all empirical distributions on a vector in $\mathcal{X}^n$ (i.e. types [16, Sec. 2], [17]) is denoted by $\mathcal{P}_n(\mathcal{X})$. The type of a vector $\boldsymbol{x}$ is denoted by $\hat{P}_{\boldsymbol{x}}(\cdot)$. For a given $Q \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(Q)$ is defined to be the set of sequences in $\mathcal{X}^n$ with type Q.

II. Statement of Main Result

Let $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output alphabets respectively. The probability of receiving a given output sequence $\boldsymbol{y}$ given that $\boldsymbol{x}$ is transmitted is given by $W^n(\boldsymbol{y}|\boldsymbol{x}) \triangleq \prod_{i=1}^n W(y_i|x_i)$. A codebook $\mathcal{C}=\{\boldsymbol{x}^{(1)},...,\boldsymbol{x}^{(M)}\}$ is known at both the encoder and decoder. The encoder receives as input a message m uniformly distributed on the set $\{1,...,M\}$, and transmits the corresponding codeword $\boldsymbol{x}^{(m)}$. Given $\boldsymbol{y}$, the decoder forms the estimate

$$\hat{m} = \operatorname*{arg\,max}_{j\in\{1,\dots,M\}} q^n(\boldsymbol{x}^{(j)},\boldsymbol{y}), \tag{1}$$

where n is the block length, and $q^n(\boldsymbol{x},\boldsymbol{y}) \triangleq \prod_{i=1}^n q(x_i,y_i)$. The function $q(x,y)$ is called the *decoding metric*, and is assumed to be non-negative and such that

$$q(x,y)=0 \iff W(y|x)=0. \tag{2}$$

In the case of a tie, a random codeword achieving the maximum in (1) is selected. In the case that $q(x,y)=W(y|x)$, i.e. maximum-likelihood (ML) decoding, the decoding rule in (1) is optimal. Otherwise, this setting is that of *mismatched decoding* [10]–[14].

We study the random-coding error probability under i.i.d. random coding, where the $M = e^{nR}$ codewords are generated independently according to

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = Q^n(\boldsymbol{x}) \triangleq \prod_{i=1}^{n} Q(x_i), \tag{3}$$

and where Q is an arbitrary input distribution. The random-coding error probability is denoted by $\overline{p}_e$.

It was shown in [12] that $\overline{p}_e \le e^{-nE_r(Q,R)}$, where

$$E_r(Q,R) \triangleq \max_{\rho\in[0,1]} E_0(Q,\rho) - \rho R \tag{4}$$

$$E_0(Q,\rho) \triangleq \sup_{s\ge 0} -\log \mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X},Y)^s \,|\, Y\big]}{q(X,Y)^s}\right)^{\rho}\right] \tag{5}$$

with $(X,Y,\overline{X}) \sim Q(x)W(y|x)Q(\overline{x})$. We showed in [18] that this exponent is tight with respect to the ensemble average for i.i.d. random coding, i.e. $\lim_{n\to\infty} -\frac{1}{n}\log \overline{p}_e = E_r$. The corresponding achievable rate is given by

$$I_{\mathrm{GMI}}(Q) \triangleq \sup_{s\ge 0} \mathbb{E}\left[\log \frac{q(X,Y)^s}{\mathbb{E}\big[q(\overline{X},Y)^s \,|\, Y\big]}\right], \tag{6}$$

which is commonly referred to as the generalized mutual information (GMI) [12]. Under ML decoding, i.e. $q(x,y) = W(y|x)$, E_r equals the exponent of Fano and Gallager [1], [2], and $I_{\mathrm{GMI}}(Q)$ equals the mutual information. The corresponding optimal choices of s in (5)–(6) are respectively given by $s = \frac{1}{1+\rho}$ and $s = 1$.

We define $\hat{\rho}(Q,R)$ to be the value of ρ achieving the maximum in (4) at rate R. From the analysis of Gallager [2, Sec. 5.6], we know that $\hat{\rho}$ equals one for all rates between 0 and some critical rate,

$$R_{\mathrm{cr}}(Q) \triangleq \max\big\{R \,:\, \hat{\rho}(Q,R) = 1\big\}, \tag{7}$$

and is strictly decreasing for all rates between $R_{\mathrm{cr}}(Q)$ and $I_{\mathrm{GMI}}(Q)$.

Similarly to [7], we define the following notion of regularity. We introduce the set

$$\mathcal{Y}_1 \triangleq \Big\{ y \,:\, q(x,y) \ne q(\overline{x},y) \text{ for some } x,\overline{x} \text{ such that } Q(x)Q(\overline{x})W(y|x)W(y|\overline{x}) > 0 \Big\} \tag{8}$$

and define (W,q,Q) to be *regular* if

$$\mathcal{Y}_1 \ne \emptyset. \tag{9}$$

When $q(x,y) = W(y|x)$, this is the *feasibility decoding is suboptimal* (FDIS) condition of [7]. We say that (W,q,Q) is *irregular* if it is not regular. A notable example of the irregular case is the binary erasure channel (BEC) under ML decoding.

Theorem 1. *Fix any (W,q) satisfying (2), input distribution Q and rate $R < I_{\mathrm{GMI}}(Q)$. The random-coding error probability for the i.i.d. ensemble in (3) satisfies*

$$\overline{p}_e \le \alpha(n,R) e^{-nE_r(Q,R)} \tag{10}$$

for sufficiently large n, where $\alpha(n,R)$ is defined as follows. If (W,q,Q) is regular, then

$$\alpha(n,R) \triangleq \begin{cases} \frac{K}{n^{\frac{1}{2}(1+\hat{\rho}(Q,R))}} & R \in \big(R_{\mathrm{cr}}(Q), I_{\mathrm{GMI}}(Q)\big) \\ \frac{K}{\sqrt{n}} & R \in \big[0, R_{\mathrm{cr}}(Q)\big], \end{cases} \tag{11}$$

and if (W,q,Q) is irregular, then

$$\alpha(n,R) \triangleq \begin{cases} \frac{K}{\sqrt{n}} & R \in \big(R_{\mathrm{cr}}(Q), I_{\mathrm{GMI}}(Q)\big) \\ 1 & R \in \big[0, R_{\mathrm{cr}}(Q)\big], \end{cases} \tag{12}$$

where K is a constant depending only on W, q, Q and R.

Proof: See Section III. ∎

In the case of ML decoding, Theorem 1 coincides with the main results of Altuğ and Wagner [7], [8] in both the regular and irregular case. Neither [7], [8] nor the present paper attempt to explicitly characterize or bound the constant K in (11)–(12). Asymptotic bounds with the constant factor specified are derived in [14] using saddlepoint approximations; see also [6] for rates below the critical rate, and [5] for strongly symmetric channels.[1]

III. Proof of Theorem 1

For a fixed value of $s \ge 0$, we define the *generalized information density* [18], [19]

$$i_s(x,y) \triangleq \log \frac{q(x,y)^s}{\sum_{\overline{x}} Q(\overline{x}) q(\overline{x},y)^s} \tag{13}$$

and its multi-letter extension

$$i_s^n(\boldsymbol{x},\boldsymbol{y}) \triangleq \sum_{i=1}^{n} i_s(x_i,y_i). \tag{14}$$

Our analysis is based on the random-coding union (RCU) bound for mismatched decoding, given by [18], [19]

$$\overline{p}_e \le \mathbb{E}\Big[\min\Big\{1, (M-1) \times \mathbb{P}\big[i_s^n(\overline{\boldsymbol{X}},\boldsymbol{Y}) \ge i_s^n(\boldsymbol{X},\boldsymbol{Y}) \,|\, \boldsymbol{X},\boldsymbol{Y}\big]\Big\}\Big], \tag{15}$$

where $(\boldsymbol{X},\boldsymbol{Y},\overline{\boldsymbol{X}}) \sim P_{\boldsymbol{X}}(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})P_{\boldsymbol{X}}(\overline{\boldsymbol{x}})$. Furthermore, we will make use of the identity

$$E_0(Q,\rho) = \sup_{s\ge 0} -\log \mathbb{E}\big[e^{-\rho i_s(X,Y)}\big] \tag{16}$$

with $(X,Y) \sim Q \times W$, which follows from (5) and (13).

We provide a number of preliminary results in Section III-A. The proof of Theorem 1 for the regular case is given in Section III-B, and the changes required to handle the irregular case are given in Section III-C.

[1] The English translation of [5] incorrectly states that the prefactor is $O\big(n^{-\frac{1}{2(1+\hat{\rho}(R))}}\big)$ for the regular case with $R > R_{\mathrm{cr}}$ (see (1.28)–(1.32) therein), but this error is not present in the original Russian version.

A. Preliminary Results

The main tool used in the proof of Theorem 1 is the following lemma by Polyanskiy *et al.* [19], which can be proved using the Berry-Esseen theorem.

Lemma 1. [19, Lemma 47] *Let $Z_1,...,Z_n$ be independent random variables with $\sigma^2 = \sum_{i=1}^n \mathrm{Var}[Z_i] > 0$ and $T = \sum_{i=1}^n \mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^3] < \infty$. Then for any t,*

$$\mathbb{E}\bigg[\exp\Big(-\sum_i Z_i\Big)\mathbb{1}\Big\{\sum_i Z_i > t\Big\}\bigg] \le 2\Big(\frac{\log 2}{\sqrt{2\pi}} + \frac{12T}{\sigma^2}\Big)\frac{1}{\sigma}\exp\big(-t\big). \quad (17)$$

The following lemma shows that under the assumption (2), we do not need to consider s growing unbounded in (5).

Lemma 2. *For any (W,q) satisfying (2), and any $\rho \in [0,1]$, the supremum in (5) is achieved (possibly non-uniquely) by some finite $s \ge 0$.*

Proof: We treat the regular and irregular cases separately. In the regular case, let $(x,\overline{x},y)$ satisfy the condition in the definition of $\mathcal{Y}_1$ in (8), and assume without loss of generality that $q(\overline{x},y) > q(x,y)$. We can upper bound the objective in (5) by

$$-\log Q(x)W(y|x)\bigg(Q(\overline{x})\Big(\frac{q(\overline{x},y)}{q(x,y)}\Big)^s\bigg)^\rho, \quad (18)$$

which tends to $-\infty$ as $s \to \infty$. It follows that the supremum is achieved by a finite value of s.

In the irregular case, we have $q(x,y) = q(\overline{x},y)$ wherever $Q(x)Q(\overline{x})W(y|x)q(\overline{x},y) > 0$, where the replacement of $W(y|\overline{x})$ by $q(\overline{x},y)$ in the latter condition follows from (2). In this case, writing the objective in (5) as

$$-\log\sum_{x,y} Q(x)W(y|x)\bigg(\sum_{\overline{x}} Q(\overline{x})\Big(\frac{q(\overline{x},y)}{q(x,y)}\Big)^s\bigg)^\rho, \quad (19)$$

we see that all choices of $s > 0$ are equivalent, since the argument to $(\cdot)^s$ equals one for all $(x,\overline{x},y)$ yielding non-zero terms in the summations. ■

The following lemma is somewhat more technical, and ensures the existence of a sufficiently high probability set in which Lemma 1 can be applied to the inner probability in (29) with a value of σ having $\sqrt{n}$ growth. We make use of the conditional distributions

$$V_s(x|y) \triangleq \frac{Q(x)q(x,y)^s}{\sum_{\overline{x}} Q(\overline{x})q(\overline{x},y)^s} \quad (20)$$

$$V_s^n(\boldsymbol{x}|\boldsymbol{y}) \triangleq \prod_{i=1}^n V_s(x_i|y_i), \quad (21)$$

which yield $i_s(x,y) = \log\frac{V_s(x|y)}{Q(x)}$ and $i_s^n(\boldsymbol{x},\boldsymbol{y}) = \log\frac{V_s^n(\boldsymbol{x}|\boldsymbol{y})}{Q^n(\boldsymbol{x})}$ (see (13)–(14)). Furthermore, we define the random variables

$$(X,Y,\overline{X},X_s) \sim Q(x)W(y|x)Q(\overline{x})V_s(x_s|y)$$
$$(\boldsymbol{X},\boldsymbol{Y},\overline{\boldsymbol{X}},\boldsymbol{X}_s) \sim Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})Q^n(\overline{\boldsymbol{x}})V_s^n(\boldsymbol{x}_s|\boldsymbol{y}), \quad (22)$$

and we write the empirical distribution of $\boldsymbol{y}$ as $\hat{P}_{\boldsymbol{y}}(\cdot)$.

Lemma 3. *If (W,q,Q) is regular and (2) holds, then the set*

$$\mathcal{F}_{n,\delta} \triangleq \Big\{\boldsymbol{y} \,:\, \sum_{y\in\mathcal{Y}_1}\hat{P}_{\boldsymbol{y}}(y) > \delta\Big\} \quad (23)$$

satisfies the following properties:

1) *For any $\boldsymbol{y} \in \mathcal{F}_{n,\delta}$, we have*

$$\mathrm{Var}\big[i_s^n(\boldsymbol{X}_s,\boldsymbol{Y})\,|\,\boldsymbol{Y}=\boldsymbol{y}\big] \ge n\delta v_s, \quad (24)$$

where

$$v_s \triangleq \min_{y\in\mathcal{Y}_1}\mathrm{Var}\big[i_s(X_s,Y)\,|\,Y=y\big]. \quad (25)$$

Furthermore, $v_s > 0$ for all $s > 0$.

2) *For all $R < I_{\mathrm{GMI}}(Q)$, there exists a choice of $\delta > 0$ such that under i.i.d. random coding,*

$$\mathbb{P}\big[\mathrm{error} \cap \boldsymbol{Y} \notin \mathcal{F}_{n,\delta}\big] \le e^{-n(E_r'(Q,R)+o(1))} \quad (26)$$

for some $E_r'(Q,R) > E_r(Q,R)$.

Proof: See the Appendix. ■

B. Proof for the Regular Case

Using the second part of Lemma 3 with the suitably chosen value of δ, and using the fact that $\lim_{n\to\infty} -\frac{1}{n}\log\overline{p}_e = E_r$ [18], we can write the random-coding error probability as

$$\overline{p}_e = \mathbb{P}\big[\mathrm{error} \cap \boldsymbol{Y} \in \mathcal{F}_{n,\delta}\big] + \mathbb{P}\big[\mathrm{error} \cap \boldsymbol{Y} \notin \mathcal{F}_{n,\delta}\big] \quad (27)$$
$$= \big(1+o(1)\big)\mathbb{P}\big[\mathrm{error} \cap \boldsymbol{Y} \in \mathcal{F}_{n,\delta}\big]. \quad (28)$$

Writing K_1 in place of $1+o(1)$ and modifying the RCU bound in (15) to include the condition $\boldsymbol{Y} \in \mathcal{F}_{n,\delta}$ in (28), we obtain

$$\overline{p}_e \le K_1 \sum_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{F}_{n,\delta}} P_{\boldsymbol{X}}(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x}) \times \min\Big\{1, M\mathbb{P}\big[i_s^n(\overline{\boldsymbol{X}},\boldsymbol{y}) \ge i_s^n(\boldsymbol{x},\boldsymbol{y})\big]\Big\}. \quad (29)$$

The value of $s \ge 0$ in (29) is arbitrary, and we choose it to achieve the supremum in (5) at $\rho = \hat{\rho}(Q,R)$, in accordance with Lemma 2. We can assume that $s > 0$, since $s = 0$ yields an objective of zero in (5), contradicting the assumption that $R < I_{\mathrm{GMI}}$.

In order to make the inner probability in (29) more amenable to an application of Lemma 1, we follow [20, Sec. 3.4.5] and write

$$Q^n(\overline{\boldsymbol{x}}) = Q^n(\overline{\boldsymbol{x}})\frac{V_s^n(\overline{\boldsymbol{x}}|\boldsymbol{y})}{V_s^n(\overline{\boldsymbol{x}}|\boldsymbol{y})} \quad (30)$$
$$= V_s^n(\overline{\boldsymbol{x}}|\boldsymbol{y})\exp\big(-i_s^n(\overline{\boldsymbol{x}},\boldsymbol{y})\big). \quad (31)$$

For a fixed sequence $\boldsymbol{y}$ and a constant t, summing both sides of (31) over all $\overline{\boldsymbol{x}}$ such that $i_s^n(\overline{\boldsymbol{x}},\boldsymbol{y}) \ge t$ yields

$$\mathbb{P}\big[i_s^n(\overline{\boldsymbol{X}},\boldsymbol{y}) \ge t\big] = \mathbb{E}\Big[\exp\big(-i_s^n(\boldsymbol{X}_s,\boldsymbol{Y})\big)\mathbb{1}\big\{i_s^n(\boldsymbol{X}_s,\boldsymbol{Y}) \ge t\big\}\,\Big|\,\boldsymbol{Y}=\boldsymbol{y}\Big] \quad (32)$$

under the joint distribution in (22). Applying Lemma 1 to (32) and using the first part of Lemma 3, we obtain for all $\boldsymbol{y} \in \mathcal{F}_{n,\delta}$ that

$$\mathbb{E}\Big[\exp\big(-i_s^n(\boldsymbol{X}_s,\boldsymbol{Y})\big)\mathbb{1}\big\{i_s^n(\boldsymbol{X}_s,\boldsymbol{Y}) \geq t\big\}\,\Big|\,\boldsymbol{Y}=\boldsymbol{y}\Big] \leq \frac{K_2}{\sqrt{n}}e^{-t} \quad (33)$$

for some constant K_2. Here we have used the fact that T in (17) grows linearly in n, which follows from the fact that we are considering finite alphabets [19, Lemma 46]. Substituting (33) into (29), we obtain

$$\overline{p}_e \leq K_1 \sum_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{F}_{n,\delta}} P_{\boldsymbol{X}}(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x}) \quad (34)$$

$$\times \min\left\{1, \frac{MK_2}{\sqrt{n}}e^{-i_s^n(\boldsymbol{x},\boldsymbol{y})}\right\} \quad (35)$$

$$\leq K_1\mathbb{E}\left[\min\left\{1, \frac{MK_2}{\sqrt{n}}e^{-i_s^n(\boldsymbol{X},\boldsymbol{Y})}\right\}\right] \quad (36)$$

$$\leq K_3\mathbb{E}\left[\min\left\{1, \frac{M}{\sqrt{n}}e^{-i_s^n(\boldsymbol{X},\boldsymbol{Y})}\right\}\right] \quad (37)$$

where (36) follows by upper bounding the summation over $\boldsymbol{y} \in \mathcal{F}_{n,\delta}$ by a summation over all $\boldsymbol{y}$, and (37) follows by defining $K_3 \triangleq K_1 \max\{1, K_2\}$.

We immediately obtain the desired result for rates below the critical rate by upper bounding the $\min\{1,\cdot\}$ term in (37) by one and using (16) (with $\rho = 1$) and the definition of i_s^n. In the remainder of the subsection, we focus on rates above the critical rate.

For any non-negative random variable A, we have $\mathbb{E}[\min\{1, A\}] = \mathbb{P}[A \geq U]$, where U is uniform on $(0, 1)$ and independent of A. We can thus write (37) as

$$\overline{p}_e \leq K_3\mathbb{P}\left[\frac{M}{\sqrt{n}}e^{-i_s^n(\boldsymbol{X},\boldsymbol{Y})} \geq U\right] \quad (38)$$

$$= K_3\mathbb{P}\left[\sum_{i=1}^n \big(R - i_s(X_i,Y_i)\big) \geq \log\big(U\sqrt{n}\big)\right]. \quad (39)$$

Let $F(t)$ denote the cumulative distribution function (CDF) of $R - i_s(X,Y)$ with $(X,Y) \sim Q\times W$, and let $Z_1,\cdots,Z_n$ be i.i.d. according to the tilted CDF

$$F_Z(z) = e^{E_r(Q,R)}\int_{-\infty}^{z} e^{\hat{\rho}t}dF(t), \quad (40)$$

where $\hat{\rho} = \hat{\rho}(Q,R)$. It is easily seen that this is indeed a CDF by writing

$$\int_{-\infty}^{\infty} e^{\hat{\rho}t}dF(t) = \mathbb{E}\big[e^{\hat{\rho}(R-i_s(X,Y))}\big] = e^{-E_r(Q,R)}, \quad (41)$$

where the last equality follows from (16) and since we have assumed that s is chosen optimally.

Similarly to [21, Lemma 2], we can use (40) to write the probability in (39) as follows:

$$\mathbb{P}\left[\sum_{i=1}^n \big(R - i_s(X_i,Y_i)\big) \geq \log\big(U\sqrt{n}\big)\right]$$

$$= \int\!\cdots\!\int_{\sum_i t_i \geq \log(u\sqrt{n})} dF(t_1)\cdots dF(t_n)dF_U(u) \quad (42)$$

$$= e^{-nE_r(Q,R)}\int\!\cdots\!\int_{\sum_i z_i \geq \log(u\sqrt{n})} e^{-\hat{\rho}\sum_i z_i} \times dF_Z(z_1)\cdots dF_Z(z_n)dF_U(u), \quad (43)$$

where $F_U(u)$ denotes the CDF of U. Substituting (43) into (39), we obtain

$$\overline{p}_e \leq K_3 e^{-nE_r(Q,R)} \times \mathbb{E}\Big[e^{-\hat{\rho}\sum_i Z_i}\mathbb{1}\Big\{\hat{\rho}\sum_i Z_i \geq \hat{\rho}\log\big(U\sqrt{n}\big)\Big\}\Big]. \quad (44)$$

Let $E_0(Q,\rho,s)$ be defined as in (5) with a fixed value of s in place of the supremum. The moment generating function (MGF) of Z is given by

$$M_Z(\tau) = \mathbb{E}[e^{\tau Z}] \quad (45)$$

$$= e^{E_r(Q,R)}\mathbb{E}\big[e^{(\hat{\rho}+\tau)(R-i_s(X,Y))}\big] \quad (46)$$

$$= e^{E_0(Q,\hat{\rho},s)}e^{-(E_0(Q,\hat{\rho}+\tau,s)-\tau R)}, \quad (47)$$

where (46) follows from (40), and (47) follows from (4) and (16). Using the identities $\mathbb{E}[Z] = \frac{dM_Z}{d\tau}\Big|_{\tau=0}$ and $\mathrm{Var}[Z] = \frac{d^2M_Z}{d\tau^2}\Big|_{\tau=0}$, we obtain

$$\mathbb{E}[Z] = R - \frac{\partial E_0(Q,\rho,s)}{\partial\rho}\Big|_{\rho=\hat{\rho}} = 0 \quad (48)$$

$$\mathrm{Var}[Z] = -\frac{\partial^2 E_0(Q,\rho,s)}{\partial\rho^2}\Big|_{\rho=\hat{\rho}} > 0, \quad (49)$$

where the second equality in (48) and the inequality in (49) hold since $R \in \big(R_{\mathrm{cr}}(Q), I_{\mathrm{GMI}}(Q)\big)$ and hence $\hat{\rho} \in (0,1)$ (e.g. see [2, pp. 142-143]). Writing the expectation in (44) as a nested expectation given U and applying Lemma 1, it follows that

$$\overline{p}_e \leq K_4 e^{-nE_r(Q,R)}\mathbb{E}\left[\frac{1}{\sqrt{n}}e^{-\hat{\rho}\log(U\sqrt{n})}\right] \quad (50)$$

$$= K_4 e^{-nE_r(Q,R)}\mathbb{E}\left[\frac{1}{\sqrt{n}}\left(\frac{1}{U\sqrt{n}}\right)^{\hat{\rho}}\right] \quad (51)$$

$$= \frac{K_4}{n^{\frac{1}{2}(1+\hat{\rho})}}e^{-nE_r(Q,R)}\mathbb{E}\big[U^{-\hat{\rho}}\big] \quad (52)$$

$$= \frac{K_5}{n^{\frac{1}{2}(1+\hat{\rho})}}e^{-nE_r(Q,R)}, \quad (53)$$

where K_4 and $K_5 = K_4\mathbb{E}\big[U^{-\hat{\rho}}\big]$ are constants. This concludes the proof.

C. Proof for the Irregular Case

The upper bound of one at rates below the critical rate in (12) was given by Kaplan and Shamai [12], so we focus on rates above the critical rate. The proof for the regular case used two applications of Lemma 1; see (33) and (50). The former leads to a multiplicative $n^{-\frac{\hat{\rho}(R)}{2}}$ term in the final expression, and the second leads to a multiplicative $n^{-\frac{1}{2}}$ term. In the irregular case, we only perform the latter application of Lemma 1. The proof is otherwise essentially identical. Applying Markov's inequality to the RCU bound in (15), we obtain

$$\overline{p}_e \le \mathbb{E}\left[\min\left\{1, Me^{-i_s^n(\boldsymbol{X},\boldsymbol{Y})}\right\}\right]. \tag{54}$$

Repeating the analysis of the regular case starting from (37), we obtain the desired result.

Appendix

Here we provide the proof of Lemma 3. The first property is easily proved by writing

$$\begin{aligned}
&\mathrm{Var}[i_s^n(\boldsymbol{X}_s,\boldsymbol{Y})\,|\,\boldsymbol{Y}=\boldsymbol{y}] && (55)\\
&\quad= \sum_{i=1}^{n}\mathrm{Var}[i_s(X_{s,i},Y_i)\,|\,Y_i=y_i] && (56)\\
&\quad\ge \sum_{y\in\mathcal{Y}_1} n\hat{P}_{\boldsymbol{y}}(y)\mathrm{Var}[i_s(X_s,Y)\,|\,Y=y]. && (57)
\end{aligned}$$

Substituting the bound on $\hat{P}_{\boldsymbol{y}}(y)$ in (23) and the definition of v_s in (25), we obtain (24). To prove that $v_s > 0$, we note that the variance of a random variable is zero if and only if the variable is deterministic, and hence

$$\begin{aligned}
&\mathrm{Var}[i_s(X_s,Y)\,|\,Y=y]=0\\
&\iff \log\frac{V_s(x|y)}{Q(x)} \text{ is independent of } x \text{ wherever } V_s(x|y)>0 && (58)\\
&\iff \frac{q(x,y)^s}{\sum_{\overline{x}}Q(\overline{x})q(\overline{x},y)^s} \text{ is independent of } x \text{ wherever } Q(x)q(x,y)^s>0 && (59)\\
&\iff q(x,y) \text{ is independent of } x \text{ wherever } Q(x)q(x,y)>0 && (60)\\
&\iff y\notin\mathcal{Y}_1, && (61)
\end{aligned}$$

where (59) follows from the definition of V_s in (20), (60) follows from the assumption $s > 0$, and (61) follows from (2) and the definition of $\mathcal{Y}_1$ in (8).

We now turn to the proof of the second property. Modifying the RCU bound in (15) to include the condition $\boldsymbol{Y}\notin\mathcal{F}_{n,\delta}$ in (26), we have for any $s\ge 0$ that

$$\begin{aligned}
&\mathbb{P}\big[\text{error} \cap \boldsymbol{Y}\notin\mathcal{F}_{n,\delta}\big] && (62)\\
&\quad\le \sum_{\boldsymbol{x},\boldsymbol{y}\notin\mathcal{F}_{n,\delta}} Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x}) \times \min\Big\{1, M\mathbb{P}\big[i_s^n(\overline{\boldsymbol{X}},\boldsymbol{y})\ge i_s^n(\boldsymbol{x},\boldsymbol{y})\big]\Big\} && (63)\\
&\quad\le \sum_{\boldsymbol{x},\boldsymbol{y}\notin\mathcal{F}_{n,\delta}} Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})\Big(Me^{-i_s^n(\boldsymbol{x},\boldsymbol{y})}\Big)^{\rho} && (64)
\end{aligned}$$

where (64) follows from Markov's inequality and since $\min\{1,\alpha\}\le\alpha^\rho$ $(0\le\rho\le 1)$. We henceforth choose ρ and s to achieve the maximum and supremum in (4) and (5) respectively, in accordance with Lemma 2. With these choices, we have similarly to (16) that

$$e^{-nE_r(Q,R)} = \sum_{\boldsymbol{x},\boldsymbol{y}} Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})\Big(Me^{-i_s^n(\boldsymbol{x},\boldsymbol{y})}\Big)^{\rho}. \tag{65}$$

Hence, we will complete the proof by showing that

$$\sum_{\boldsymbol{x},\boldsymbol{y}\notin\mathcal{F}_{n,\delta}} Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})e^{-\rho i_s^n(\boldsymbol{x},\boldsymbol{y})} \tag{66}$$

has a strictly larger exponential rate of decay than

$$\sum_{\boldsymbol{x},\boldsymbol{y}} Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})e^{-\rho i_s^n(\boldsymbol{x},\boldsymbol{y})} \tag{67}$$

for some $\delta > 0$. By performing an expansion in terms of types, (67) is equal to

$$\begin{aligned}
&\sum_{P_{XY}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})} \mathbb{P}\big[(\boldsymbol{X},\boldsymbol{Y})\in T^n(P_{XY})\big]e^{-n\rho\mathbb{E}_P[i_s(X,Y)]} && (68)\\
&\quad= \max_{P_{XY}}\exp\Big(-n\big(D(P_{XY}\|Q\times W) + \rho\mathbb{E}_P[i_s(X,Y)]+o(1)\big)\Big), && (69)
\end{aligned}$$

where (69) follows from the property of types in [17, Eq. (12)] and the fact that the number of joint types is polynomial in n. Substituting the definitions of divergence and i_s (see (13)) into (69), we see that the exponent of (67) equals

$$\min_{P_{XY}}\sum_{x,y}P_{XY}(x,y) \times \log\left(\frac{P_{XY}(x,y)}{Q(x)W(y|x)}\left(\frac{q(x,y)^s}{\sum_{\overline{x}}Q(\overline{x})q(\overline{x},y)^s}\right)^{\rho}\right). \tag{70}$$

Similarly, and from the definition of $\mathcal{F}_{n,\delta}$ in (23), (66) has an exponent equal to

$$\min_{P_{XY}\,:\,\sum_{y\in\mathcal{Y}_1}P_Y(y)\le\delta}\sum_{x,y}P_{XY}(x,y) \times \log\left(\frac{P_{XY}(x,y)}{Q(x)W(y|x)}\left(\frac{q(x,y)^s}{\sum_{\overline{x}}Q(\overline{x})q(\overline{x},y)^s}\right)^{\rho}\right). \tag{71}$$

A straightforward evaluation of the Karush-Kuhn-Tucker (KKT) conditions [22, Sec. 5.5.3] yields that (70) is uniquely minimized by

$$P_{XY}^{*}(x,y) = \frac{Q(x)W(y|x)\left(\frac{\sum_{\overline{x}} Q(\overline{x})q(\overline{x},y)^s}{q(x,y)^s}\right)^{\rho}}{\sum_{x',y'} Q(x')W(y'|x')\left(\frac{\sum_{\overline{x}'} Q(\overline{x}')q(\overline{x}',y')^s}{q(x',y')^s}\right)^{\rho}}. \quad (72)$$

From the assumptions in (2) and (9), we can find a symbol $y^* \in \mathcal{Y}_1$ such that $P_Y^*(y^*) > 0$. Choosing $\delta < P_Y^*(y^*)$, it follows that P_{XY}^* fails to satisfy the constraint in (71), and thus (71) is strictly greater than (70).

REFERENCES

[1] R. Fano, *Transmission of information: A statistical theory of communications*. MIT Press, 1961.

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

[3] C. E. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[4] P. Elias, "Coding for two noisy channels," in *Third London Symp. Inf. Theory*, 1955.

[5] R. L. Dobrushin, "Asymptotic estimates of the probability of error for transmission of messages over a discrete memoryless communication channel with a symmetric transition probability matrix," *Theory Prob. Appl.*, vol. 7, no. e, pp. 270–300, 1962.

[6] 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.

[7] Y. Altuğ and A. B. Wagner, "A refinement of the random coding bound," in *50th Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.

[8] ——, "Refinement of the random coding bound," in *Int. Zurich. Sem. Comms.*, Zurich, Feb. 2012.

[9] ——, "Refinement of the sphere-packing bound," in *IEEE Int. Symp. Inf. Theory*, 2012, pp. 2949–2953.

[10] 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.

[11] 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.

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

[13] 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.

[14] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Finite-length bounds, error exponents and approximations," submitted to *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1303.6166].

[15] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements and connections," submitted To *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1307.6679].

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

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

[18] 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.

[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] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, 2010.

[21] R. Bahadur and R. Ranga Rao, "On deviations of the sample mean," *The Annals of Mathematical Statistics*, vol. 31, pp. 1015–1027, Dec. 1960.

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