Concentration of Random-Coding Error Exponents

Lan V. Truong
University of Cambridge
lt407@cam.ac.uk

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

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

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

***Abstract*—This paper studies the error exponent of i.i.d. randomly generated codes used for transmission over discrete memoryless channels with maximum likelihood decoding. Specifically, this paper shows that the error exponent of a code, defined as the negative normalized logarithm of the probability of error, converges in probability to the typical error exponent. For high rates, the result is a consequence of the fact that the random-coding error exponent and the sphere-packing error exponent coincide. For low rates, instead, the proof of convergence is based on the fact that the union bound accurately characterizes the probability of error.**

I. Introduction

In [1], Shannon used the i.i.d. random-coding ensemble to show that for discrete memoryless channels (DMC) there exist codes whose probability of error vanishes for rates below the channel capacity. For the same rate regime, Fano [2] characterized the exponential decay of the error probability defining the error exponent as the negative normalized logarithm of the ensemble-average error probability. In [3], Gallager derived the error exponent for DMC in a simpler way while improving at low rates using the idea of expurgation. A lower bound on the error probability in DMC, called sphere-packing bound, was first introduced in [4], and Nakiboglu [5] recently derived sphere-packing bounds for some stationary memoryless channels using Augustin's method. The corresponding sphere-packing exponent is shown to coincide with the random-coding exponent for rates higher than a certain critical rate.

Barg and Forney [6] derived the typical random-coding error exponent (TRC) for the random-coding ensemble over the binary symmetric channel. Upper and lower bounds on the TRC for fixed-constant composition codes and general discrete memoryless channels were provided in [7]. For the same type of codes and channels, Merhav [8] determined the exact TRC error exponent for a wide class of stochastic decoders called generalized likelihood decoder (GLD), maximum-likelihood being a particular case. Merhav derived the TRC exponent for spherical codes over colored Gaussian channels [9] and for random convolutional code ensembles [10]. The error exponent of a random fixed-composition code with GLD is known to converge in probability to the TRC [11], a convergence that is non-symmetric: the lower tail decays exponentially while the upper tail decays doubly-exponentially. The latter was first established for a limited range of rates in [12]. For pairwise-independent ensembles and arbitrary channels, Cocco et al. [13] showed that the probability that the exponent of a given code in the ensemble is smaller than a lower bound on the TRC exponent is vanishingly small. The interest in the TRC exponent lies in the fact that is the largest exponent that can be achieved for a given ensemble. This is in contrast with the expurgated exponent, as codes that attain it no longer belong to a pairwise independent ensemble [3], [14].

II. Prelimilaries

We consider the problem of transmitting information reliably over a DMC with transition probability W and respective finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. In particular, we study the transmission of M_n equiprobable messages using a code $\mathcal{c}_n$ with codewords $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{M_n}$, $\boldsymbol{x}_i \in \mathcal{X}^n$, $\forall i = 1, \dots, M_n$. The error probability of such code is

$$P_{\mathrm{e}}(\mathcal{c}_n) = \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbb{P}\Bigg[\bigcup_{j \neq i} \{\boldsymbol{x}_i \to \boldsymbol{x}_j\} \Bigg], \tag{1}$$

where $\{\boldsymbol{x}_i \to \boldsymbol{x}_j\}$ is the pairwise error event under maximum likelihood decoding, i.e., the event of deciding in favor of codeword $\boldsymbol{x}_j$ when codeword $\boldsymbol{x}_i$ was transmitted. The error exponent of $\mathcal{c}_n$ is defined as

$$E_n(\mathcal{c}_n) = -\frac{1}{n} \log P_{\mathrm{e}}(\mathcal{c}_n). \tag{2}$$

Let $R = \lim_{n\to\infty} \frac{1}{n} \log M_n$ be the rate of the code in bits per channel use. An error exponent $E(R)$ is achievable when there exists a sequence of codes $\{\mathcal{c}_n\}_{n=1}^{\infty}$ such that $\liminf_{n\to\infty} E_n(\mathcal{c}_n) \geq E(R)$.

We next consider the i.i.d. random-coding ensemble, the set of codes $\mathcal{C}_n$ whose codewords $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{M_n}$ are pairwise-independently generated with a single-letter input distribution Q. Similarly to random variables, $\mathcal{C}_n$ denotes a random code, and $\mathcal{c}_n$ denotes a specific code in the ensemble. The random-coding error exponent $E_{\mathrm{rce}}(R)$ as originally derived by Fano and Gallager is defined as

$$E_{\mathrm{rce}}(R) = \lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)], \tag{3}$$

where the expectation is taken over the code ensemble [3]. Observe that (3) suggests that $E_{\mathrm{rce}}(R)$ is the asymptotic exponent of the ensemble-average probability of error. Instead,

This work has been funded in part by the European Research Council under ERC grant agreement 725411, by the Secretary of Universities and Research (Catalan Government) under a Beatriu de Pinós postdoctoral fellowship, and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 801370.

the typical random-coding exponent is defined as the limiting expected error exponent over the ensemble, that is,

$$E_{\mathrm{trc}}(R) = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log P_{\mathrm{e}}(\mathcal{C}_n)\big]. \tag{4}$$

III. Main Result

We next state our main result using the notion of convergence in probability. A sequence of random variables $\{A_n\}_{n=1}^{\infty}$ converges to A in probability, denoted as $A_n \xrightarrow{\mathrm{p}} A$ if for all $\delta > 0$, $\lim_{n\to\infty}\mathbb{P}[|A_n - A| > \delta] = 0$ [15, Sec. 2.2]. The following result states the concentration of the error exponent $E_n(\mathcal{C}_n)$ to the TRC exponent $E_{\mathrm{trc}}(R)$ for rates below the channel capacity C.

Theorem 1. *For a general DMC channel, for all rates R such that $0 \le R \le C$, it holds that*

$$E_n(\mathcal{C}_n) \xrightarrow{\mathrm{p}} E_{\mathrm{trc}}(R). \tag{5}$$

Before proceeding with some details of the proof, we discuss some of the implications of the above result. Theorem 1 not only proves the achievability of the TRC exponent, but also shows that the probability of finding a code in the ensemble with higher or lower exponent than the TRC exponent tends to zero. The above concentration property gives more information about the error exponent behavior of the ensemble than the traditional derivation of the random coding error exponent, which computes the exponent of the expected error probabiltiy. This way, the TRC emerges as the most likely error exponent in the random-coding ensemble as the block length n tends to infinity —if one wishes to improve the error exponent, one must improve the ensemble. The work in [11] shows such a concentration property by separately studying the the tails of the distribution of $E_n(\mathcal{C}_n)$ for the constant composition ensemble and DMCs. It shows an interesting asymmetry: the probability $\mathbb{P}[E_n(\mathcal{C}_n) < E_{\mathrm{trc}}(R)]$ decays exponentially, while $\mathbb{P}[E_n(\mathcal{C}_n) > E_{\mathrm{trc}}(R)]$ decays double-exponentially. This implies that, beyond the concentration property, it is significantly more difficult to find a code in the ensemble with exponent higher than $E_{\mathrm{trc}}(R)$. This asymmetry is difficult to obtain from the proof of Theorem 1, as one would need to study separately the two tails, as done in [11].

As we shall see, the proof of Theorem 1 requires different techniques for the rate regimes $0 \le R < R_{\mathrm{crit}}(Q)$ and $R_{\mathrm{crit}}(Q) \le R \le C$, where $R_{\mathrm{crit}}(Q)$ is the critical rate in [3, Eq. (36)], that is the rate above which the sphere packing exponent $E_{\mathrm{sp}}(R)$ coincides with the random-coding exponent $E_{\mathrm{rce}}(R)$ [16, Sec. 5.8]. Thus, in this rate region, we can expect that $E_{\mathrm{trc}}(R) = E_{\mathrm{rce}}(R)$.

IV. Proof of Theorem 1

In standard concentration inequalities, sequences of random variables are usually assumed to be independent, group-independent, or dependent on each other according to known structures [17]. Here, we are dealing with the concentration of the probability of error $P_{\mathrm{e}}(\mathcal{C}_n)$ in (1). This expression is a sum of dependent random variables where each term in the sum contains all random vectors $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_M\}$ and channel noise. Thus, this expression does not belong to the aforementioned structures that enable the use of standard concentration inequalities. While $P_{\mathrm{e}}(\mathcal{C}_n)$ can be considered as a function of independent random variables, this function is only in implicit form. It is therefore not simple to obtain accurate bounds on the variance, or even show that these bounds on the variance tend to zero, by directly using the standard concentration inequalities such as the Efron-Stein inequality [18]. In this paper, we develop an easier way of upper bounding the variance, which makes use of various existing results.

For the range of rate $R_{\mathrm{crit}}(Q) \le R \le C$, the proof is based on Levy's continuity theorem [19] and benefits from the fact that in this rate regime, the following equalities hold

$$E_{\mathrm{sp}}(R) = E_{\mathrm{rce}}(R) = E_{\mathrm{trc}}(R). \tag{6}$$

In particular, we consider the moment-generating function of the negative error exponent, given by

$$\phi(\lambda) = \mathbb{E}\big[2^{\lambda\frac{\log P_{\mathrm{e}}(\mathcal{C}_n)}{n}}\big]. \tag{7}$$

Using the sphere-packing bound [4] to the error probability in (7), we obtain that

$$\phi(\lambda) \ge 2^{-\lambda E_{\mathrm{sp}}(R)}. \tag{8}$$

Similarly, since $P_{\mathrm{e}}(\mathscr{c}_n)^{\lambda/n}$ is a concave function for the range $0 \le \lambda \le n$, Jensen's inequality implies that

$$\phi(\lambda) \le 2^{-\lambda E_{\mathrm{rce}}(R)}. \tag{9}$$

Hence, using (6), we obtain that $\phi(\lambda)$ converges to the moment-generating function of the constant $-E_{\mathrm{trc}}(R)$.

The rest of the paper is devoted to the proof of the main result for the range of rates $0 \le R < R_{\mathrm{crit}}(Q)$. We first need some definitions and lemmas. For this range, the proof uses the union bound to the error probability (1), that is

$$P_{\mathrm{e}}(\mathscr{c}_n) \le P_{\mathrm{e}}^{\mathrm{ub}}(\mathscr{c}_n), \tag{10}$$

where $P_{\mathrm{e}}^{\mathrm{ub}}(\mathscr{c}_n)$ is given by

$$P_{\mathrm{e}}^{\mathrm{ub}}(\mathscr{c}_n) = \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j], \tag{11}$$

and we define its finite-length error exponent as

$$E_n^{\mathrm{ub}}(\mathscr{c}_n) = -\frac{1}{n}\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathscr{c}_n). \tag{12}$$

Under i.i.d. random coding, the pairwise error probabilities $\mathbb{P}[\boldsymbol{X}_i \to \boldsymbol{X}_j]$, with, $i \neq j$ are pairwise-independent random variables. Hence, the union bound $P_{e}^{\mathrm{ub}}(\mathcal{C}_n)$ in (11) is a sum of M_n pairwise-independent random variables, where concentration properties are expected to hold [11]. Let $E_{\mathrm{trc}}^{\mathrm{ub}}(R)$ be the union-bound TRC exponent of (11), that is

$$E_{\mathrm{trc}}^{\mathrm{ub}}(R) = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)\big]. \tag{13}$$

The next result shows that the probability that the union-bound exponent (12) deviates from the union-bound TRC exponent (13) vanishes as $n \to \infty$.

Lemma 2. *For all rate R such that $0 \leq R < R_{\text{crit}}(Q)$, any $\varepsilon > 0$ and for some $\kappa > 0$, it holds that*

$$\mathbb{P}\left[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) > \frac{1}{2}2^{-n(E_{\text{trc}}^{\text{ub}}(R)-\varepsilon)}\right] + \mathbb{P}\left[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) < 2^{-n(E_{\text{trc}}^{\text{ub}}(R)+\varepsilon)}\right] \leq \frac{1}{n^{1+\kappa}}. \quad (14)$$

It then remains to relate the convergence in probability of the original error exponent (2) to that of the union bound (12). To do so, we use the following Lemma that is based on Caen's inequality [20].

Lemma 3. *For all rate R such that $0 < R < R_{\text{crit}}(Q)$ and for some $\delta(R) > 0$, it holds that*

$$0 \leq \frac{\mathbb{E}[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n)]}{\mathbb{E}[P_{\text{e}}(\mathcal{C}_n)]} - 1 \leq 2^{-n\left(\delta(R)+E_{\text{trc}}^{\text{ub}}(R)-E_{\text{rce}}(R)\right)}. \quad (15)$$

We are now equipped to prove Theorem 1 by observing that for any $\varepsilon > 0$, the convergence in probability of $E_n(\mathcal{C}_n)$ to $E_{\text{trc}}(R)$ can be written and upper bounded as

$$\begin{aligned}\mathbb{P}\big[|E_n(\mathcal{C}_n) - E_{\text{trc}}(R)| > 3\varepsilon\big] &\leq \underbrace{\mathbb{P}\Big[\big|E_n(\mathcal{C}_n) - E_n^{\text{ub}}(\mathcal{C}_n)\big| > \varepsilon\Big]}_{\alpha_n} \\ &+ \underbrace{\mathbb{P}\left[\left|E_n^{\text{ub}}(\mathcal{C}_n) - \left(-\frac{1}{n}\mathbb{E}\big[\log P_{\text{e}}^{\text{ub}}(\mathcal{C}_n)\big]\right)\right| > \varepsilon\right]}_{\beta_n} \\ &+ \underbrace{\mathbb{P}\left[\left|\left(-\frac{1}{n}\mathbb{E}\big[\log P_{\text{e}}^{\text{ub}}(\mathcal{C}_n)\big]\right) - E_{\text{trc}}(R)\right| > \varepsilon\right]}_{\gamma_n}. \quad (16)\end{aligned}$$

We next show that the terms α_n, β_n and γ_n in (16) tend to zero as $n \to \infty$, implying the concentration result in (5).

A. First term of (16)

The term α_n quantifies the deviation of the error exponent of the error probability (2) with that of the union bound (11). By the symmetry of the i.i.d. random-coding ensemble, for any pair of codewords $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ with $i \neq j$ we have that

$$\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_i \to \boldsymbol{X}_j]\big] = \mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_1 \to \boldsymbol{X}_2]\big]. \quad (17)$$

Similarly, for any triplet of codewords $\boldsymbol{X}_i$, $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$ with $j, k \neq i$ and $j \neq k$, it holds that

$$\mathbb{E}\Big[\mathbb{P}\big[\{\boldsymbol{X}_i \to \boldsymbol{X}_j\} \cap \{\boldsymbol{X}_i \to \boldsymbol{X}_k\}\big]\Big] = \mathbb{E}\Big[\mathbb{P}\big[\{\boldsymbol{X}_1 \to \boldsymbol{X}_2\} \cap \{\boldsymbol{X}_1 \to \boldsymbol{X}_3\}\big]\Big]. \quad (18)$$

In both (17) and (18), the expectations are calculated with respect to the i.i.d. ensemble codeword distribution $Q^n(\boldsymbol{x}) = \prod_{k=1}^n Q(x_k)$, where $Q(x)$ is the single-letter input distribution. We next provide separate convergence of α_n for $R = 0$ and for $0 < R < R_{\text{crit}}(Q)$. For the simple case of $R = 0$, we first observe that the union bound (11) can be bounded as

$$\begin{aligned}P_{\text{e}}^{\text{ub}}(\mathcal{c}_n) &= \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j] && (19)\\ &\leq (M_n - 1)\max_{i\neq j}\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j], && (20)\end{aligned}$$

while the probability of error (1) can be lower bounded by

$$\begin{aligned}P_{\text{e}}(\mathcal{c}_n) &= \frac{1}{M_n}\sum_{i=1}^{M_n}\mathbb{P}\Big[\bigcup_{j\neq i}\{\boldsymbol{x}_i \to \boldsymbol{x}_j\}\Big] && (21)\\ &\geq \frac{1}{M_n}\max_{i\neq j}\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j]. && (22)\end{aligned}$$

From (20) and (22), we have that the first term in the r.h.s. of (16), after some algebra, satisfies

$$\begin{aligned}\alpha_n &= \mathbb{P}\Big[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) > 2^{n\varepsilon}P_{\text{e}}(\mathcal{C}_n)\Big] && (23)\\ &\leq \mathbb{P}\Big[(M_n - 1)\max_{i\neq j}\mathbb{P}[\boldsymbol{X}_i \to \boldsymbol{X}_j] > 2^{n\varepsilon}\frac{1}{M_n}\max_{i\neq j}\mathbb{P}[\boldsymbol{X}_i \to \boldsymbol{X}_j]\Big] && (24)\\ &= \mathbb{P}\Big[(M_n - 1) > 2^{n\varepsilon}\frac{1}{M_n}\Big]. && (25)\end{aligned}$$

Since M_n is any sub-exponential sequence in n, the probability in (25) vanishes as $n \to \infty$ for $\varepsilon > 0$.

We now consider the case of $0 < R < R_{\text{crit}}(Q)$. We define the sequence a_n as

$$a_n = 2^{-n(E_{\text{trc}}^{\text{ub}}(R)+\frac{\varepsilon}{2})}. \quad (26)$$

Then, the first term in the r.h.s. of (16) can be written, after some algebra, as

$$\begin{aligned}\alpha_n &= \mathbb{P}\Big[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) - a_n - 2^{n\varepsilon}\big(P_{\text{e}}(\mathcal{C}_n) - a_n\big) > (2^{n\varepsilon} - 1)a_n\Big] && (27)\\ &\leq \mathbb{P}\Big[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) - a_n > \frac{1}{2}(2^{n\varepsilon} - 1)a_n\Big] + \mathbb{P}\Big[-2^{n\varepsilon}\big(P_{\text{e}}(\mathcal{C}_n) - a_n\big) > \frac{1}{2}(2^{n\varepsilon} - 1)a_n\Big], && (28)\end{aligned}$$

where (28) follows from the fact that for any three random variables A, B and C, the tail probability $\mathbb{P}[A + B > 2C]$ satisfies $\mathbb{P}[A + B > 2C] \leq \mathbb{P}[A > C] + \mathbb{P}[B > C]$. We next bound the two terms in the r.h.s. of (28) using the definition of a_n in (26). For the first term, we have

$$\begin{aligned}\mathbb{P}\Big[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) - a_n > \frac{1}{2}(2^{n\varepsilon} - 1)a_n\Big] &= \mathbb{P}\Big[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) > \frac{1}{2}(2^{n\varepsilon} + 1)a_n\Big] && (29)\\ &= \mathbb{P}\Big[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) > \frac{1}{2}(2^{n\varepsilon} + 1)2^{-n(E_{\text{trc}}^{\text{ub}}(R)+\frac{\varepsilon}{2})}\Big] && (30)\\ &\leq \mathbb{P}\Big[P_{\text{e}}^{\text{ub}}(\mathcal{C}_n) > \frac{1}{2}2^{-n(E_{\text{trc}}^{\text{ub}}(R)-\frac{\varepsilon}{2})}\Big], && (31)\end{aligned}$$

whereas for the second term we have

$$\begin{aligned}
&\mathbb{P}\Big[-2^{n\varepsilon}\big(P_{\mathrm{e}}(\mathcal{C}_n)-a_n\big) > \frac{1}{2}(2^{n\varepsilon}-1)a_n\Big] \\
&= \mathbb{P}\Big[2^{n\varepsilon}\big(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-P_{\mathrm{e}}(\mathcal{C}_n)\big) \\
&\qquad -2^{n\varepsilon}\big(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-a_n\big) > \frac{1}{2}(2^{n\varepsilon}-1)a_n\Big] && (32)\\
&\le \mathbb{P}\Big[2^{n\varepsilon}\big(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-P_{\mathrm{e}}(\mathcal{C}_n)\big) > \frac{1}{4}(2^{n\varepsilon}-1)a_n\Big] \\
&\quad + \mathbb{P}\Big[-2^{n\varepsilon}\big(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-a_n\big) > \frac{1}{4}(2^{n\varepsilon}-1)a_n\Big]. && (33)
\end{aligned}$$

We proceed by bounding the second term of (33) as

$$\begin{aligned}
&\mathbb{P}\Big[-2^{n\varepsilon}\big(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-a_n\big) > \frac{1}{4}(2^{n\varepsilon}-1)a_n\Big] \\
&= \mathbb{P}\Big[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) < \Big(1-\frac{1}{4}\Big(\frac{2^{n\varepsilon}-1}{2^{n\varepsilon}}\Big)\Big)a_n\Big] && (34)\\
&\le \mathbb{P}\Big[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}^{\mathrm{ub}}(R)+\frac{\varepsilon}{2})}\Big], && (35)
\end{aligned}$$

while the first term of (33) can be bounded[1] as

$$\begin{aligned}
&\mathbb{P}\Big[2^{n\varepsilon}\big(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-P_{\mathrm{e}}(\mathcal{C}_n)\big) > \frac{1}{4}(2^{n\varepsilon}-1)a_n\Big] \\
&\stackrel{.}{\le} a_n^{-1}\mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-P_{\mathrm{e}}(\mathcal{C}_n)] && (36)\\
&= 2^{(E_{\mathrm{trc}}^{\mathrm{ub}}(R)+\frac{\varepsilon}{2})n}\mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-P_{\mathrm{e}}(\mathcal{C}_n)], && (37)
\end{aligned}$$

where (36) follows from the union bound in (10) and Markov's inequality, and (37) follows from the definition of a_n in (26). Next, from Lemma 3, we have the exponential upper bound

$$\begin{aligned}
&\mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-P_{\mathrm{e}}(\mathcal{C}_n)] \\
&\quad = \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\left(\frac{\mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)]}{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}-1\right) && (38)\\
&\quad \stackrel{.}{\le} 2^{-nE_{\mathrm{rce}}(R)}2^{-n\left(\delta(R)+E_{\mathrm{trc}}^{\mathrm{ub}}(R)-E_{\mathrm{rce}}(R)\right)}. && (39)
\end{aligned}$$

Using (39) back in (37) and simplifying the result, we obtain

$$\mathbb{P}\Big[2^{n\varepsilon}\big(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)-P_{\mathrm{e}}(\mathcal{C}_n)\big) > \frac{a_n}{4}(2^{n\varepsilon}-1)\Big] \stackrel{.}{\le} 2^{-n\left(\delta(R)-\frac{\varepsilon}{2}\right)}. \tag{40}$$

Hence, from (33), (35), and (40), we have that the second term of (28) is upper bounded as

$$\begin{aligned}
&\mathbb{P}\Big[-2^{n\varepsilon}\big(P_{\mathrm{e}}(\mathcal{C}_n)-a_n\big) > \frac{1}{2}(2^{n\varepsilon}-1)a_n\Big] \\
&\stackrel{.}{\le} \mathbb{P}\Big[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}^{\mathrm{ub}}(R)+\frac{\varepsilon}{2})}\Big] + 2^{-n\left(\delta(R)-\frac{\varepsilon}{2}\right)}. && (41)
\end{aligned}$$

Finally, combining (31) and (41) in (28) and using Lemma 2, we obtain that the first term of (16) satisfies

$$\alpha_n \stackrel{.}{\le} \frac{1}{n^{1+\kappa}} + 2^{-n\left(\delta(R)-\frac{\varepsilon}{2}\right)} \tag{42}$$

Since α_n in (16) is non-increasing with ε, it holds that $\lim_{n\to\infty}\alpha_n = 0$ and therefore that $E_n(\mathcal{C}_n)\xrightarrow{\mathrm{p}}E_{\mathrm{trc}}(R)$. Using the Borel-Cantelli lemma with (42) and the dominated covergence theorem [21], we obtain that the TRC error exponent is the same as the estimate obtained from the union bound, i.e., $E_{\mathrm{trc}}(R) = E_{\mathrm{trc}}^{\mathrm{ub}}(R)$.

B. Second term of (16)

Using Chebyshev's inequality, we have

$$\beta_n \le \frac{\sigma_n^2}{\varepsilon^2}, \tag{43}$$

where σ_n^2 is the variance of the random variable $E_n^{\mathrm{ub}}(\mathcal{C}_n)$. Using (11)–(12), we have that

$$\begin{aligned}
\sigma_n^2 &= \frac{1}{n^2}\mathbb{E}\left[\left(-\log(M_n-1)-\log\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{M_n-1}\right)\right)^2\right] \\
&\quad - \left(\frac{\mathbb{E}\big[-\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)\big]}{n}\right)^2 && (44)\\
&= \frac{1}{n^2}\mathbb{E}\left[\left(-\log(M_n-1)-\log\xi_n\right.\right. \\
&\quad \left.\left.-\log\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{(M_n-1)\xi_n}\right)\right)^2\right] - \left(\frac{\mathbb{E}\big[-\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)\big]}{n}\right)^2, && (45)
\end{aligned}$$

where in (45) we introduced the variable ξ_n defined as

$$\xi_n = 2^{-n(E_{\mathrm{trc}}^{\mathrm{ub}}(R)+R)}. \tag{46}$$

Using (45) and the definition (13), we obtain that (43) satisfies

$$\begin{aligned}
\limsup_{n\to\infty}\beta_n &\le \limsup_{n\to\infty}\frac{1}{\varepsilon^2}\mathbb{E}\left[\left(E_{\mathrm{trc}}^{\mathrm{ub}}(R)\right.\right. \\
&\quad \left.\left.-\frac{1}{n}\log\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{(M_n-1)\xi_n}\right)\right)^2\right] - \frac{E_{\mathrm{trc}}^{\mathrm{ub}}(R)^2}{\varepsilon^2}. && (47)
\end{aligned}$$

Expanding squares, we further obtain

$$\begin{aligned}
\limsup_{n\to\infty}\beta_n &\le \frac{1}{\varepsilon^2}\bigg\{E_{\mathrm{trc}}^{\mathrm{ub}}(R)^2 - 2E_{\mathrm{trc}}^{\mathrm{ub}}(R) \\
&\quad \times \liminf_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\log\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{(M_n-1)\xi_n}\right)\right] \\
&\quad + \limsup_{n\to\infty}\mathbb{E}\left[\left(\frac{1}{n}\log\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{(M_n-1)\xi_n}\right)\right)^2\right]\bigg\} \\
&\quad - \frac{(E_{\mathrm{trc}}^{\mathrm{ub}}(R)^2)}{\varepsilon^2}. && (48)
\end{aligned}$$

By using the bounded convergence theorem and the continuous mapping theorem [21], we can show that the expectation terms in (48) vanish as $n\to\infty$, that is

$$\mathbb{E}\left[\frac{1}{n}\log\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{(M_n-1)\xi_n}\right)\right] \to 0, \tag{49}$$

$$\mathbb{E}\left[\left(\frac{1}{n}\log\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{(M_n-1)\xi_n}\right)\right)^2\right] \to 0. \tag{50}$$

From (48), (49), and (50), we conclude that $\lim_{n\to\infty}\beta_n = 0$ for any arbitrary $\varepsilon > 0$.

[1] We write $a_n \stackrel{.}{\le} b_n$ for two positive sequences $\{a_n\}_{n\ge1}$ and $\{b_n\}_{n\ge1}$ such that $\frac{1}{n}\log\frac{a_n}{b_n} \le 0$.

C. Third term of (16)

The analysis of the first term of (16), in Section IV-A implies that $E_{\mathrm{trc}}^{\mathrm{ub}}(R) = E_{\mathrm{trc}}(R)$. Therefore, the third term of (16) also vanishes for any $\varepsilon > 0$ because $-\frac{1}{n}\mathbb{E}\big[\log P_e^{\mathrm{ub}}(\mathcal{C}_n)\big] \to E_{\mathrm{trc}}^{\mathrm{ub}}(R)$.

In conclusion, as anticipated, the three terms of (16) tend to zero as $n \to \infty$, showing (5) for rates below the critical rate. Together with (6)–(9), we proved Theorem 1, the convergence in probability of the error exponent of codes in the ensemble to the typical random-coding error exponent.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[2] R. M. Fano, *Transmission of Information*. New York: Wiley, 1961.

[3] R. G. Gallager, "Simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Th.*, vol. 11, no. 3, pp. 3–18, Jan 2008.

[4] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding in discrete memoryless channels I-II," *Information and Control*, vol. 10, pp. 65–103, 522–552, 1967.

[5] B. Nakiboglu, "The sphere packing bound for memoryless channels," *Problems of Information Transmission*, vol. 56, pp. 201–244, 2020.

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

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

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

[9] ——, "Error exponents of typical random codes for the colored gaussian channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8164–8179, 2019.

[10] ——, "Error exponents of typical random trellis codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2067–2077, 2019.

[11] 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.

[12] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 430–443, 1982.

[13] G. Cocco, A. Guillén i Fàbregas, and J. Font-Segura, "A dual-domain achievability of the typical error exponent," in *IEEE Int. Symp. Inf. Th.*, Melbourne, Australia, 2021.

[14] 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.

[15] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge Univ. Press, 2010.

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

[17] G. L. S. Boucheron and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[18] G. Lugosi, "Concentration of measure inequalities," *Lecture Notes, Universitat Pompeu Fabra*, 2021.

[19] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. John Wiley and Sons, 1971.

[20] D. de Caen, "A lower bound on the probability of a union," *Discrete Math.*, vol. 69, pp. 217–220, May 1997.

[21] P. Billingsley, *Probability and Measure*, 3rd ed. Wiley-Interscience, 1995.