Convergence in Distribution of the Error Exponent of Random Codes at Zero Rate

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

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

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

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

***Abstract*—We study the convergence in distribution of the error exponent of random codes, defined as the negative normalized logarithm of the probability of error, of both i.i.d. and constant-composition ensembles over discrete memoryless channels. For a constant number of messages, the distribution of the error exponent converges to that of the minimum of a set of independent normal random variables. For an increasing sub-exponential number of messages, the error exponent converges to a normal distribution, independent of the number of messages. As a by-product, we provide a new method to prove the convergence to a normal distribution of an infinite number of random variables based on a modification of the Wasserstein metric.**

I. Introduction

We consider reliable information transmission over a discrete memoryless channel (DMC) with transition probability W, and finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. We study the transmission of M_n equiprobable messages using a code $c_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{M_n}\}$, where each codeword $\boldsymbol{x}_i \in \mathcal{X}^n$. The channel output is denoted by $\boldsymbol{y} \in \mathcal{Y}^n$. The error probability of such code is given by

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

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

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

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 said to be achievable if there exists a sequence of codes $\{c_n\}_{n=1}^{\infty}$ such that $\liminf_{n\to\infty} E_n(c_n) \geq E(R)$.

We define the i.i.d. and constant composition random-coding ensembles as the set of codes $\mathcal{C}_n$ whose codewords $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{M_n}$ are independently generated with either a single-letter input distribution Q or equiprobably from all sequences with the same empirical distribution. Let $\mathcal{T}_n(\mathcal{X}\times\mathcal{X})$ be the set of all n-joint types on $\mathcal{X} \times \mathcal{X}$. An n-type Q_X is called regular if

$$|\mathcal{T}_n(\mathcal{X}\times\mathcal{X})| \left(\frac{|\mathcal{T}(Q_X)|^2 - |\mathcal{T}(Q^*_{XX'})|}{|\mathcal{T}(Q_X)|^2} \right) \to 0 \tag{3}$$

as $n \to \infty$, where $Q^*_{XX'} := Q_X Q_X$. The regular condition of types assumes that the rate of convergence to zero of $\Pr[(\boldsymbol{X}_i, \boldsymbol{X}_j) \notin \mathcal{T}(Q^*_{XX'})]$ is faster than $O(1/(n+1)^{|\mathcal{X}|^2})$

Similarly to random variables (RVs), $\mathcal{C}_n$ denotes a random code, and c_n denotes a specific code in the ensemble. The random-coding error exponent $E_{\mathrm{rce}}(R)$ is defined as

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

where the expectation is taken over the code ensemble [1], [2]. Observe that (4) suggests that $E_{\mathrm{rce}}(R)$ is the asymptotic exponent of the ensemble-average probability of error. Instead, the typical random-coding exponent is defined as the limiting expected error exponent over the ensemble, that is,

$$E_{\mathrm{trc}}(R) = \lim_{n\to\infty} -\frac{1}{n} \mathbb{E}\big[\log P_{\mathrm{e}}(\mathcal{C}_n) \big]. \tag{5}$$

The typical error exponent $E_{\mathrm{trc}}(R)$ has been shown to be significantly larger than the random-coding exponent $E_{\mathrm{rce}}(R)$ for some channels and code ensembles in the region of low rates [3]–[6]. Zero-rate transmission has applications in areas such as machine-to-machine or Covert communications.

In our previous work [7], we studied the convergence in probability of the error exponent to the typical random-coding exponent for the i.i.d. code ensemble over DMCs, that is

$$E_n(\mathcal{C}_n) \xrightarrow{(\mathrm{p})} E_{\mathrm{trc}}(R), \tag{6}$$

for $0 \leq R < C$, where $A_n \xrightarrow{(\mathrm{p})} A$ is a placeholder for the condition $\lim_{n\to\infty} \mathbb{P}[|A_n - A| > \delta] = 0$ [8, Sec. 2.2], for all $\delta > 0$. The convergence in probability for constant composition codes over DMCs was shown in [9]. Here, we introduce results related to the convergence in distribution of the error exponent $E_n(\mathcal{C}_n)$, valid for both i.i.d. and constant-composition ensembles at asymptotically zero rate. A sequence of RVs $\{A_n\}_{n=1}^{\infty}$ converges to A in distribution, denoted as $A_n \xrightarrow{(\mathrm{d})} A$ if $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \big|\mathbb{P}[A_n \leq x] - \mathbb{P}[A \leq x]\big| = 0$ [8, Sec. 2.2] for all points of continuity x of $\mathbb{P}[A \leq x]$.

This work has been funded by the European Research Council under ERC grant 725411, by the Catalan Government under a Beatriu de Pinós fellowship, and by the Marie Skłodowska-Curie programme under grant 801370.

II. Main Results

The following results state the concentration in distribution of the normalized error exponent $E_n(\mathcal{C}_n)$ as $n \to \infty$, for a constant number of messages and for an increasing sub-exponential number of messages.

Theorem 1: Let $U_{ij} \sim \mathcal{N}(0,1)$, for $i = 1, \ldots, M$ and $j = 1, \ldots, M$ such that $i \neq j$, be a set of independent standard normal RVs. Then, the error exponent of random i.i.d. and constant-composition codes (under the regular condition (3)) with a constant number of codewords, i.e., $M_n = M$ for some constant M, satisfies

$$\frac{E_n(\mathcal{C}_n) - \mathbb{E}[E_n(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(E_n(\mathcal{C}_n))}} \xrightarrow{\text{(d)}} \frac{\min_{i\neq j} U_{ij} - \mathbb{E}[\min_{i\neq j} U_{ij}]}{\sqrt{\mathrm{Var}(\min_{i\neq j} U_{ij})}}. \quad (7)$$

The proof of Theorem 1, sketched in Sec. III, is based on the fact that $E_n(\mathcal{C}_n)$ is a minimization of a constant number of $M(M-1)$ terms where each term is a sum of independent RVs in the i.i.d. ensemble, and a sum of dependent RVs with an additional vanishing term in the constant-composition ensemble. In Fig. 1, we plot the histogram of the error exponent $E_n(\mathcal{C}_n)$ for both the i.i.d. and constant-composition ensembles over a binary symmetric channel (BSC) with crossover probability $p = 0.11$, input distribution $Q(x) = \frac{1}{2}$, $M = 4$ codewords and length $n = 10,000$, after 10^7 trials. For comparison, we also plot the asymptotic distribution of the random variable $\min_{i\neq j} U_{ij}$ in the right-hand side of (7) (solid), and a normal approximation with the same mean and variance (dashed). We note that the error exponent converges to a Gaussian-like distribution, with a slightly asymmetric tilting in both tails, also observed in the histogram.

Theorem 2: Let M_n be an increasing subexponential number of messages, namely $\lim_{n\to\infty} M_n = \infty$ and $\lim_{n\to\infty} (\log M_n)/n = 0$, but growing fast enough such that

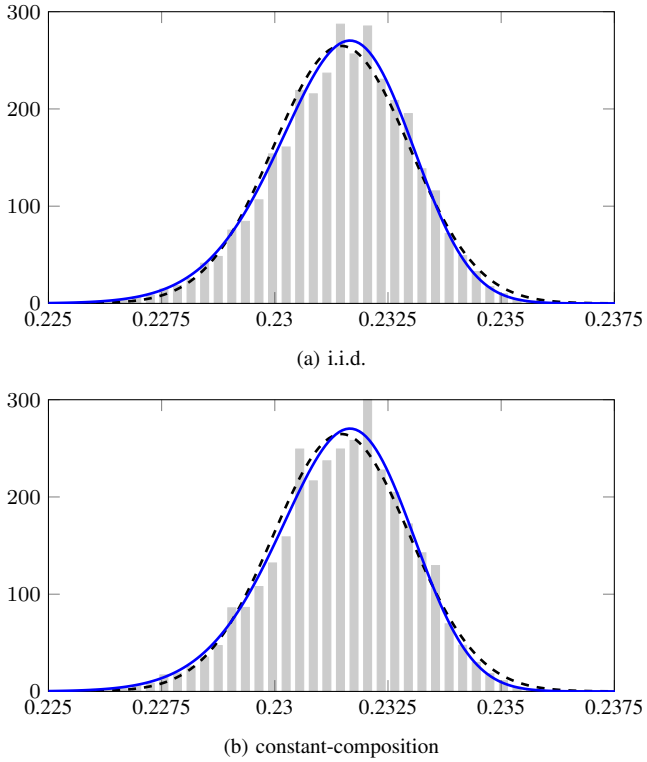
$$\sum_{n=1}^{\infty} \frac{1}{M_n(M_n - 1)} < \infty. \quad (8)$$

Then, the error exponent of random i.i.d. and constant-composition codes (under the regular condition (3)) satisfies

$$\frac{E_n(\mathcal{C}_n) - \mathbb{E}[E_n(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(E_n(\mathcal{C}_n))}} \xrightarrow{\text{(d)}} \mathcal{N}(0,1). \quad (9)$$

The proof of Theorem 2, sketched in Sec. IV, is based on the fact that, under the condition $\lim_{n\to\infty} (\log M_n)/n = 0$, for both i.i.d. and constant-composition, the error exponent $E_n(\mathcal{C}_n)$ is the minimum of an infinite number of RVs, each converging to a Gaussian distribution (see Eq. (18) below).

For a constant number of messages $M_n = M$, the condition in Theorem 2 is not satisfied, and therefore the error exponent does not concentrate according to (9) but to (7) instead. The fact that M_n grows with n implies that the dependence between the codeword symbols vanishes as $n \to \infty$, and therefore the independence of U_{ij} is preserved. On the contrary, for a (sub-exponentially) growing number of messages M_n, the dependence among the codewords, and therefore the correlation among U_{ij}, increases such that the RVs U_{ij} can be represented by a common Gaussian random variable U. One example of sub-exponential growth of the number of messages satisfying (8) is the polynomial function of n given, for some $\delta > 0$, by $M_n = \Omega\big(n^{\frac{1+\delta}{2}}\big)$.



(a) i.i.d.

(b) constant-composition

Fig. 1. Distribution of the error exponent of the (a) i.i.d. and (b) constant-composition codes over the BSC with $M = 4$, $n = 10,000$, symmetric input distribution and composition, and $p = 0.11$. Histograms of $E_n(\mathcal{C}_n)$ with 10^7 trials, dashed black lines are normal distributions, and solid blue lines are the distributions of $\min_{i\neq j} U_{ij}$.

III. Proof of Theorem 1

We start with the proof of the convergence in distribution for a constant number of messages. The proof is based on the Stein's method for the convergence in distribution of the sum of RVs to a Gaussian random variable. For the i.i.d. ensemble, this is a direct application of the central limit theorem (CLT). For the constant-composition ensemble, we modify the Stein's criteria in [10, Theorem 3.2] to accommodate for the dependence among the RVs in the following lemma. A detailed proof can be found in [11, Appendix L].

Lemma 1: Let $X_1, X_2, \cdots, X_n$ be zero-mean RVs such that $\sum_{i=1}^n \mathbb{E}[X_i^2] = n$. Assume there exist positive sequences $\{\xi_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$, and a set $\mathcal{V} \subset \mathbb{R}^n$ such that

$$\big(1 - \xi_n\big) \prod_{i=1}^{n} \mathbb{P}[X_i = x_i] \leq$$

$$\mathbb{P}[X_1 = x_1, \cdots, X_n = x_n] \leq \big(1 + \xi_n\big) \prod_{i=1}^{n} \mathbb{P}[X_i = x_i] \quad (10)$$

for all $x_1, x_2, \cdots, x_n \in \mathcal{V}$, and such that

$$\max\left\{1, \frac{1}{\sqrt{n}}\sum_{i=1}^{n}|x_i|, \frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2, \frac{1}{n}\sum_{i=1}^{n}x_i^2, \frac{1}{n^{3/2}}\sum_{i=1}^{n}|x_i|^3\right\} \leq g_n, \quad (11)$$

for all $(x_1, x_2, \cdots, x_n) \in \mathcal{V}^c$. Assume also that

$$g_n \max\{\mathbb{P}(V^c), \mathbb{P}_{\Pi}(V^c)\} \to 0 \quad \text{as} \quad n \to \infty, \quad (12)$$

as $n \to \infty$, where we defined $\mathbb{P}_{\Pi}$ as the product probability measure, i.e., $\mathbb{P}_{\Pi}[x_1, x_2, \cdots, x_n] = \prod_{i=1}^{n}\mathbb{P}[X_i = x_i]$. Now, let $S_n = X_1 + X_2 + \cdots + X_n$ and define T as

$$T = \frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}. \quad (13)$$

Under the condition that

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}[|X_i^3|] \to 0, \quad (14)$$

$$\frac{1}{n^{2}}\sum_{i=1}^{n}\mathbb{E}[|X_i|^4] \to 0, \quad (15)$$

we have

$$T \xrightarrow{\text{(d)}} \mathcal{N}(0,1). \quad (16)$$

Lemma 1 recovers the original Stein's criterion in [10, Theorem 3.2] for independent random variables setting $\mathcal{V}^c = \emptyset$.

We are now equipped to prove Theorem 1. Bounding the union of pairwise events in the right-hand of (1), we first observe that the error probability of a code $P_{\mathrm{e}}(c_n)$ with M_n messages can be lower and upper bounded as

$$\max_{i\neq j}\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j] \leq P_{\mathrm{e}}(\mathcal{C}_n) \leq M_n(M_n-1)\max_{i\neq j}\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j]. \quad (17)$$

From (17), for any M_n sub-exponential in n, it holds that

$$E_n(\mathcal{C}_n) \sim \min_{i\neq j} Z_{ij}(n) \quad (18)$$

where $Z_{ij}(n)$ is the exponent of the pairwise error probability $\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j]$ given in terms of the Bhattacharyya distance between two symbols $d_{\mathrm{B}}(x,x') = -\log\left(\sum_y \sqrt{W(y|x)W(y|x')}\right)$ as

$$Z_{ij}(n) = -\frac{1}{n}\sum_{k=1}^{n}\sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbf{1}\{(X_{ik}, X_{jk}) = (x,x')\}. \quad (19)$$

In (19), $\mathbf{1}\{\cdot\}$ is the indicator function, and X_{ik} denotes the k-th symbol of codeword $\boldsymbol{X}_i$. Equation (18) states that the error exponent almost surely equals the minimum of a sequence of RVs $Z_{ij}(n)$, $i,j \in \{1,\ldots,M\}$, $i \neq j$, each being the sum of n RVs according to (19).

We next study the convergence of distribution of $Z_{ij}(n)$ as $n \to \infty$, defining $T_{ij}(n)$ as

$$T_{ij}(n) = \frac{Z_{ij}(n) - \mathbb{E}[Z_{ij}(n)]}{\sqrt{\mathrm{Var}(Z_{ij}(n))}}. \quad (20)$$

For the i.i.d. ensemble, $Z_{ij}(n)$ is a sum of n i.i.d. RVs, so that the CLT implies $T_{ij}(n) \xrightarrow{\text{(d)}} \mathcal{N}(0,1)$, $\forall i \neq j$. For the constant-composition ensemble, we will use the method of types to write (19) as the sum of type class RVs that satisfy the conditions of Lemma 1. Let Q_X be the codeword type and $Q_{XX'}$ the induced joint type. We have that (19) can be written as

$$Z_{ij}(n) = -\sum_{Q_{XX'}}\sum_{x,x'} Q_{XX'}(x,x')d_{\mathrm{B}}(x,x')A^{ij}_{Q_{XX'}}, \quad (21)$$

where, for a given pair of random codewords $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ we defined the random variable $A^{ij}_{Q_{XX'}}$ as

$$A^{ij}_{Q_{XX'}} = \mathbf{1}\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q_{XX'})\} \quad (22)$$

and $\mathcal{T}_n(Q_{XX'})$ is the type class. In order to write each random variable $Z_{ij}(n)$ in (21) in a convenient way to use Lemma 1, we define the random variable $U^{ij}_{Q_{XX'}}$ as

$$U^{ij}_{Q_{XX'}} = \sqrt{\frac{|\mathcal{T}_n(\mathcal{X}\times\mathcal{X})|}{\sum_{Q_{XX'}}\mathbb{E}[(V^{ij}_{Q_{XX'}})^2]}}V^{ij}_{Q_{XX'}}, \quad (23)$$

where

$$V^{ij}_{Q_{XX'}} = \sum_{x,x'} Q_{XX'}(x,x')d_{\mathrm{B}}(x,x')A^{ij}_{Q_{XX'}} - \sum_{x,x'} Q_{XX'}(x,x')d_{\mathrm{B}}(x,x')\mathbb{E}[A^{ij}_{Q_{XX'}}]. \quad (24)$$

Then,

$$\frac{Z_{ij}(n) - \mathbb{E}[Z_{ij}(n)]}{\sqrt{\mathrm{Var}(Z_{ij}(n))}} = \frac{\sum_{Q_{XX'}}U^{ij}_{Q_{XX'}}}{\sqrt{\mathrm{Var}\left(\sum_{Q_{XX'}}U^{ij}_{Q_{XX'}}\right)}}. \quad (25)$$

We note that $U^{ij}_{Q_{XX'}}$ satisfies

$$\mathbb{E}[U^{ij}_{Q_{XX'}}] = 0, \quad (26)$$

$$\sum_{Q_{XX'}}\mathbb{E}[(U^{ij}_{Q_{XX'}})^2] = |\mathcal{T}_n(\mathcal{X}\times\mathcal{X})|. \quad (27)$$

We next define a suitable set $\mathcal{V}_{ij}$ for each pair of random codewords $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ as the choice of $\mathcal{V}$ in Lemma 1, i.e.,

$$\mathcal{V}_{ij} = \Big\{\{a_{Q_{XX'}}\}_{Q_{XX'}} : \text{the only } n\text{-joint type } Q_{XX'} \text{ such that } a^*_{Q_{XX'}} = 1 \text{ is } Q^*_{XX'} = Q_X Q_X\Big\}, \quad (28)$$

where $\{a_{Q_{XX'}}\}_{Q_{XX'}}$ is the set of all possible realizations of $\{A^{ij}_{Q_{XX'}}\}_{Q_{XX'}}$ in (22). Observe that $Z_{ij}(n)$ expressed as in (21) is the sum of $|\mathcal{T}_n(\mathcal{X}\times\mathcal{X})|$ Bernoulli RVs whose joint probability distribution can be shown to be bounded as

$$\begin{aligned}(1-\xi_n)\prod_{Q_{XX'}\in\mathcal{T}_n(\mathcal{X}\times\mathcal{X})}\mathbb{P}[A^{ij}_{Q_{XX'}} = a_{Q_{XX'}}] &\leq \mathbb{P}\left[\bigcap_{Q_{XX'}\in\mathcal{T}_n(\mathcal{X}\times\mathcal{X})}\{A^{ij}_{Q_{XX'}} = a_{Q_{XX'}}\}\right] \\ &\leq (1+\xi_n)\prod_{Q_{XX'}\in\mathcal{T}_n(\mathcal{X}\times\mathcal{X})}\mathbb{P}[A^{ij}_{Q_{XX'}} = a_{Q_{XX'}}], \quad (29)\end{aligned}$$

for any sequence $\{a_{Q_{XX'}}\} \in \mathcal{V}_{ij}$ and some positive sequence $\xi_n \to 0$. Furthermore, since $V^{ij}_{Q_{XX'}}$ is bounded for all $Q_{XX'}$, we have that, in analogy to (27)

$$\sum_{Q_{XX'}} \mathbb{E}[(V^{ij}_{Q_{XX'}})^2] = \Theta\big(|\mathcal{T}_n(\mathcal{X}\times\mathcal{X})|\big). \tag{30}$$

For any sequence $\{a_{Q_{XX'}}\} \in \mathcal{V}_{ij}$, it can be shown that

$$\mathbb{P}[\{A^{ij}_{Q_{XX'}}\} \in \mathcal{V}^c_{ij}] = \frac{|\mathcal{T}(Q_X)|^2 - |\mathcal{T}(Q^*_{XX'})|}{|\mathcal{T}(Q_X)|^2} \tag{31}$$

$$\mathbb{P}_\Pi[\{A^{ij}_{Q_{XX'}}\} \in \mathcal{V}^c_{ij}] = \mathbb{P}[\{A^{ij}_{Q_{XX'}}\} \in \mathcal{V}^c_{ij}] + h_n 2^{-nI_{\min}}, \tag{32}$$

where $I_{\min}$ is a positive constant and h_n is sub-exponential in n. Since $V^{ij}_{Q_{XX'}}$ is linear in $A^{ij}_{Q_{XX'}}$ (cf. (24)), the existence of a set $\mathcal{V}$ as in Lemma 1 is guaranteed (via a linear transformation of the set $\mathcal{V}_{ij}$), i.e., $\mathbb{P}[\mathcal{V}^c] = \mathbb{P}[\{A^{ij}_{Q_{XX'}}\} \in \mathcal{V}^c_{ij}]$, $\mathbb{P}_\Pi[\mathcal{V}^c] = \mathbb{P}_\Pi[\{A^{ij}_{Q_{XX'}}\} \in \mathcal{V}^c_{ij}]$.

On the other hand, for all $\{a_{Q_{XX'}}\} \in \mathcal{V}^c_{ij}$, $V^{ij}_{Q_{XX'}}$ is bounded for all $Q_{XX'} \in \mathcal{T}_n(\mathcal{X}\times\mathcal{X})$. Hence, all the terms inside the max operator in (11) applied to the sequence of random variables $A^{ij}_{Q_{XX'}}$ are bounded such that

$$g_n = O\big(|\mathcal{T}_n(\mathcal{X}\times\mathcal{X})|\big). \tag{33}$$

From (31), (32), and (33), under the regular condition on the type Q_X given in (3), the condition (12) in Lemma 1 is attained.

Similarly, we can check all the conditions of Lemma 1 for the sum $S_n = \sum_{Q_{XX'}} U^{ij}_{Q_{XX'}}$ of $|\mathcal{T}_n(\mathcal{X}\times\mathcal{X})|$ terms. Hence, we conclude that for the constant-composition ensemble, as well as for the i.i.d. ensemble, the random variable $T_{ij}(n)$ defined in (20) converges in distribution to a normal:

$$T_{ij}(n) \xrightarrow{(\mathrm{d})} \mathcal{N}(0,1). \tag{34}$$

It only remains to relate the asymptotic distribution of $E_n(\mathcal{C}_n)$ in (18) to that of $T_{ij}(n)$ in (34). We first note that, since for any sequence of real numbers $\{\alpha_{ij}\}_{i\neq j}$, the linear combination $\sum_{i\neq j}\alpha_{ij}T_{ij}(n)$ is asymptotically Gaussian, it follows that the sequence of RVs $\{T_{ij}(n)\}_{i\neq j}$ is asymptotically jointly Gaussian for both ensembles. Finally, since $E_n(\mathcal{C}_n)$ in (18) is a bounded and continuous function of $Z_{ij}(n)$, the continuous mapping theorem [12] implies (7). See [11] for a detailed proof.

IV. Proof of Theorem 2

Unlike the constant number of messages case, the boundedness assumption used in the continuous mapping argument is not valid for a sub-exponentially increasing number of codewords. Instead, we shall consider probability metrics to measure the distance between two distributions. We start by revising the standard probability metrics and arguing that they are not refined enough for our proof.

Definition 1: Let $\mathcal{H}$ be a family of real-valued test functions $h(x)$. For two RVs X and Y, we define the probability metric $d_{\mathcal{H}}(X,Y)$ as

$$d_{\mathcal{H}}(X,Y) = \sup_{h\in\mathcal{H}} \big|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]\big|. \tag{35}$$

By properly choosing the family of test functions $h(x)$ in (35), we recover some of the standard probability metrics used in the results relative to convergence in distribution [10]. For example, taking $\mathcal{H}$ as the set of indicator functions $h(x) = \mathbf{1}\{x \le u\}$ for $u \in \mathbb{R}$, $d_{\mathcal{H}}(X,Y)$ becomes the Kolmogorov metric , while the set of functions satisfying the condition $|h(x) - h(y)| \le |x - y|$ leads to the Wasserstein metric in (35).

Bounds on the Kolmogorov and Wasserstein metrics, e.g. [10, Prop. 1.2] turn out to be not tight enough to prove the convergence in distribution to the normal random variable of the minimum of an infinite number of sums of random variables (as $n \to \infty$). We propose the following variation of the Wasserstein probability metric by modifying the probability metric (35) and further constraining the family of test functions.

Definition 2: For two RVs X and Y, we define the modified Wasserstein probability metric $\bar{d}_W(X,Y)$ as

$$\bar{d}_W(X,Y) = \sup_{h\in\mathcal{H}} \min\Big\{\big|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]\big|, \big|\mathbb{E}[h(-X)] - \mathbb{E}[h(Y)]\big|\Big\}, \tag{36}$$

where $\mathcal{H}$ is the set of test functions $h(x)$ given, for all values $a \in \mathbb{R}$ and all bounded c such that $0 < c \le 4\sqrt{2\pi}$, by

$$h(x) = \begin{cases} c & x \le a \\ 0 & x \ge a + c \\ \text{linear}, & \text{otherwise.} \end{cases} \tag{37}$$

Our modified Wasserstein metric $\bar{d}_W$ in Definition 2 allows to upper bound the absolute difference of cumulative distribution functions of a standard Gaussian random variable and any random variable as follows (see [11] for a detailed proof).

Lemma 2: For any random variable T it holds that

$$\big|\mathbb{P}[T \le x] - \mathbb{P}[Z \le x]\big| \le 2(8\pi)^{-1/4}\sqrt{\bar{d}_W(T,Z)} + \big|\mathbb{P}[T \le x] - \mathbb{P}[T \ge -x]\big|, \tag{38}$$

for all $x \in \mathbb{R}$ and where $Z \sim \mathcal{N}(0,1)$.

Inspecting Lemma 2, we note that a vanishing upper bound on $\bar{d}_W$ will imply convergence in probability of T to a standard Gaussian. In the following lemma, we obtain an upper bound to $\bar{d}_W$ that is tighter than the standard bounds to the Wasserstein metric in [10, Prop. 1.2]. By using this new bound, we can later prove the convergence in distribution to the normal random variable of the minimum of an *infinite number* of sums of of random variables where the bound of the Wasserstein metric in [10, Prop. 1.2] fails to work.

Lemma 3: For a given test function $h(x) \in \mathcal{H}$ and a random variable Z, let $f_h(x)$ be the the function that satisfies the differential equation $f'_h(x) - xf_h(x) = h(x) - \mathbb{E}[h(Z)]$. Let $T = \min\{T_1, T_2, \ldots, T_L\}$ for some $L \in \mathbb{N}$, where $(T_1, T_2, \ldots, T_L)$ is a sequence of identically distributed RVs,

and let Z be a standard Gaussian. Then, our modified Wasserstein metric in (36) is upper bounded as

$$\bar{d}_W(T,Z) \leq \mu_n + \sup_{h\in\mathcal{H}} \min\Big\{\mathbb{E}[h(T)-h(T_1)], \mathbb{E}[h(-T_1)-h(-T)]\Big\}, \quad (39)$$

where for convenience we defined μ_n as

$$\mu_n = \max\Big\{ \sup_{h\in\mathcal{H}} \big|\mathbb{E}[f_h'(T_1) - T_1 f_h(T_1)]\big|, \sup_{h\in\mathcal{H}} \big|\mathbb{E}[f_h'(-T_1) + T_1 f_h(-T_1)]\big|\Big\}. \quad (40)$$

Before starting the main part of the proof of Theorem 2, we state next that the term μ_n in the upper bound (39) is properly bounded when T_1 is a sum of independent RVs.

Lemma 4: [10, Theorem 3.2] Let $X_1, X_2, \cdots, X_n$ be a sequence of independent zero-mean RVs satisfying $\mathbb{E}[|X_i|^4] < \infty$ and $\mathbb{E}[X_i^2] = 1$. If $T_1 = \sum_{i=1}^n X_i/\sqrt{n}$ and $f_h(x)$ satisfies the conditions in Lemma 3 for a standard Gaussian Z, then the parameter μ_n in (40) is upper bounded as

$$\mu_n \leq \frac{1}{\sqrt{n^3}} \sum_{i=1}^n \mathbb{E}[|X_i|^3] + \sqrt{\frac{2}{\pi n^2} \sum_{i=1}^n \mathbb{E}[X_i^4]}. \quad (41)$$

Using the former results, we are now ready to prove Theorem 2. As in Theorem 1, for a sub-exponential number of codewords, the error exponent (2) of a randomly generated code $\mathcal{C}_n$ converges almost surely as (18), where $Z_{ij}(n)$ is again given in (19). Similarly, we define $T_{ij}(n)$ as the normalized version of $Z_{ij}(n)$ given in (20). Since the error exponent $E_n(\mathcal{C}_n)$ is related to the minimum over $i \neq j$ of $Z_{ij}(n)$, we start by studying the convergence in probability of the random variable $\min_{i\neq j} T_{ij}(n)$ as $n \to \infty$.

The arguments used to show that $T_{ij}(n)$ converges in distribution to the standard Gaussian distribution, that is (34) are also valid for a sub-exponential number of codewords. In contrast to Theorem 1, the main difficulty now is that we face a minimization of $M_n(M_n-1)$ random variables where each random variable converges in distribution to the standard normal random variable.

Let V_n be the sequence of RVs given by the normalized sum of the terms involved in $\min_{i\neq j} T_{ij}(n)$,

$$V_n = \frac{1}{M_n(M_n-1)} \sum_{i\neq j} T_{ij}(n). \quad (42)$$

and consider the probability that such sequence is bounded away from zero, that is, $\mathbb{P}[|V_n| \geq \varepsilon]$ for $\varepsilon > 0$. If condition (8) in Theorem 2 is met, then the sum of the probability of such events is finite, that is,

$$\sum_{n=1}^{\infty} \mathbb{P}[|V_n| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{M_n(M_n-1)} < \infty. \quad (43)$$

As a result of (43), the Borel–Cantelli lemma implies a vanishing probability that infinitely many V_n are away from zero, or equivalently, that almost surely (with probability one), we have $|V_n| < \varepsilon$ for all but finitely many n. Using the same arguments for the sequence of RVs given by

$$U_n = \frac{1}{M_n(M_n-1)} \sum_{i\neq j} T_{ij}(n) - T_{12}(n), \quad (44)$$

for some $k \neq l$ in $k \in \{1,\ldots,M\}$ and $l \in \{1,\ldots,M\}$, almost surely (with probability one), we have $|U_n| < \varepsilon$ for all but finitely many n. In addition, we can show that $\{T_{ij}(n) - T_{12}(n)\}_{i\neq j}$ are asymptotically pairwise independent. Combining both results, the probability that infinitely many events $\{|V_n| \geq \varepsilon\} \cap \{|U_n| \geq \varepsilon\}$ occur vanishes, or equivalently, that almost surely (with probability one), as $n \to \infty$, we have

$$-\varepsilon < V_n < \varepsilon; \qquad -\varepsilon < U_n < \varepsilon. \quad (45)$$

Using (45), the structure of the test functions (37) in Definition 2, and the bounded convergence theorem [12] as $n \to \infty$ (and $\varepsilon \to 0$), it can be shown that the second term in the right-hand side of (39), applied to the sequence of RVs $T_{ij}(n)$, vanishes as $n \to \infty$ (see [11] for a proof), i.e.:

$$\lim_{n\to\infty} \min\Big\{\mathbb{E}\Big[h\big(\min_{i\neq j} T_{ij}(n)\big) - h\big(T_{12}(n)\big)\Big], \mathbb{E}\Big[h\big(-T_{12}(n)\big) - h\big(-\min_{i\neq j} T_{ij}(n)\big)\Big]\Big\} = 0. \quad (46)$$

Using that $T_{ij}(n)$, defined in (20), is a sequence of independent zero-mean RVs, Lemma 4 implies that the parameter μ_n in the right-hand side of (39) also vanishes,

$$\lim_{n\to\infty} \mu_n = 0. \quad (47)$$

Combining (46) and (47) back in Lemma 3, we obtain that the distance between the distribution of $\min_{i\neq j} T_{ij}(n)$ and that of a standard Gaussian Z, measured with our modified Wasswestein metric in (36), vanishes, that is

$$\lim_{n\to\infty} \bar{d}_W\big(\min_{i\neq j} T_{ij}(n), Z\big) = 0. \quad (48)$$

It can be shown that the distribution of $\min_{i\neq j} T_{ij}(n)$ is tight. Thanks to this fact, we can prove that $\lim_{n\to\infty} |\mathbb{P}(\min_{i\neq j} T_{ij}(n) \leq x) - \mathbb{P}(\min_{i\neq j} T_{ij}(n) \geq -x)| = 0$ for all $x \in \mathbb{R}$ and x is a continuous point of the limiting distribution of $\min_{i\neq j} T_{ij}(n)$ (see [11, Proof of Theorem 6]). Hence, by Lemma 2, for all continuous points x in the limiting distribution of $\min_{i\neq j} T_{ij}(n)$, we have a vanishing $\lim_{n\to\infty} \big|\mathbb{P}[\min_{ij} T_{ij}(n) \leq x] - \mathbb{P}[Z \leq x]\big| = 0$ or, equivalently, convergence in distribution

$$\min_{i\neq j} T_{ij}(n) \xrightarrow{(\mathrm{d})} \mathcal{N}(0,1). \quad (49)$$

In addition, using the relations between $T_{ij}(n)$ and $Z_{ij}(n)$ in (20), it can be shown [11] that the mean and variance of $\min_{i\neq j} Z_{ij}(n)$ are related to that of Z_{12} as

$$\mathbb{E}\big[\min_{i\neq j} Z_{ij}(n)\big] = \mathbb{E}[Z_{12}] + o(1) \quad (50)$$

$$\mathrm{Var}\big(\min_{i\neq j} Z_{ij}(n)\big) = (1+o(1))\,\mathrm{Var}(Z_{12}(n)). \quad (51)$$

Finally, from $E_n(\mathcal{C}_n)$ in equations (18), and applying Slutsky's theorem [12], we obtain (9), concluding the proof.

REFERENCES

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

[2] R. G. Gallager, "Simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 3, pp. 3–18, Jan 1965.

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

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

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

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

[7] L. Truong, G. Cocco, J. Font-Segura, and A. Guillén i Fàbregas, "Concentration of random-coding error exponents," in *IEEE Information Theory Workshop (ITW)*, Kanazawa, Japan, oct 2021, pp. 1317–1321.

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

[9] 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.

[10] N. Ross, "Fundamentals of Stein's method," *Probability Surveys*, vol. 8, no. none, pp. 210 – 293, 2011.

[11] L. V. Truong, G. Cocco, J. Font-Segura, and A. Guillén i Fàbregas, "Concentration properties of random codes," *ArXiv*, vol. abs/2203.07853, 2022.

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