Concentration Properties of Random Codes

Lan V. Truong, *Member, IEEE,* Giuseppe Cocco, *Senior Member, IEEE,* Josep Font-Segura, *Member, IEEE,* and Albert Guillén i Fàbregas *Fellow, IEEE*

***Abstract*—This paper shows that, for discrete memoryless channels, the error exponent of a randomly generated code with independent codewords converges in probability to its expectation—the typical error exponent. For high rates, the result follows from the fact that the random-coding error exponent and the sphere-packing error exponent coincide. For low rates, instead, the convergence is based on the fact that the union bound accurately characterizes the error probability. The paper also zooms into the behavior at asymptotically low rates, and shows that the normalized error exponent converges in distribution to the standard Gaussian or a Gaussian-like distribution. We also state several results on the convergence of the error probability and error exponent for generic ensembles and channels.**

I. Introduction

Shannon [1] showed that for every discrete memoryless channel (DMC), there exist codes whose probability of error vanishes with the codeword length for rates below the channel capacity. Since then, significant research effort has been devoted to studying properties of the probability of error of such codes. For rates below capacity, Fano [2] characterized the exponential decay of the error probability defining the error exponent as the negative logarithm of the ensemble-average error probability normalized by the block-length, i.e., the random coding exponent (RCE). In [3], Gallager derived the RCE in a simpler way and introduced the idea of expurgation in order to obtain an improved error exponent at low rates. A lower bound on the error probability in the DMC, called sphere-packing bound, was first introduced in [4] and it was shown to coincide with the RCE for rates higher than a certain critical rate. Nakiboğlu in [5] recently derived sphere-packing bounds for some stationary memoryless channels using Augustin's method [6].

L. V. Truong is with the School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester CO4 3SQ, U.K. (e-mail: lan.truong@essex.ac.uk). G. Cocco is with the Signal Theory and Communications Department, Universitat Politècnica de Catalunya, Barcelona 08034, Spain, (e-mail: giuseppe.cocco@upc.edu). J. Font-Segura is with Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (e-mail: josep.font@upf.edu). A. Guillén i Fàbregas is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. and the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (e-mail: guillen@ieee.org).

This work has been funded in part by the European Research Council under ERC grant agreement 725411, by the Spanish Ministry of Economy and Competitiveness under Grant PID2020-116683GB-C22, by the Ramon y Cajal fellowship program (grant RYC2021-033908-I) funded by the Spanish Ministry for Science and Innovation/State Research Agency MCIN/AEI/10.13039/501100011033, by a Beatriu de Pinós postdoctoral fellowship (Catalan Government), and by the Marie Skłodowska-Curie grant agreement 801370.

This work has been presented in part at the 2021 IEEE Information Theory Workshop, Kanazawa, Japan and the 2022 IEEE Information Theory Workshop, Mumbai, India.

In [7], Barg and Forney studied the independently and identically distributed (i.i.d.) random-coding ensemble over the binary symmetric channel (BSC) with maximum likelihood decoding. They showed that the probability of finding a code with an error exponent arbitrarily close to the so-called typical random coding (TRC) exponent approaches 1 as the codeword length grows. They also showed that TRC exponent is strictly larger than the RCE at low rates, and that it coincides with the expurgated exponent at rate zero. Upper and lower bounds on the TRC for constant-composition codes and general DMCs were provided in [8]. For the same type of codes and channels, Merhav [9] determined the exact TRC error exponent for a wide class of stochastic decoders called generalized likelihood decoders (GLD), of which maximum-likelihood decoder is a special case. Merhav derived the TRC exponent for spherical codes over coloured Gaussian channels [10] and for random convolutional code ensembles [11]. Tamir *et al.* [12] studied the upper and lower tails of the error exponent around the mean, the typical error exponent, for random pairwise-independent constant-composition codes with GLD. It was shown that the tails behave in a non-symmetric way: the lower tail decays exponentially while the upper tail decays double-exponentially; the latter was first established for a limited range of rates in [13]. By studying the behavior of both tails, the work in [12] implicitly proves concentration in probability. The TRC was recently shown to be universally achievable with a likelihood mutual-information decoder in [14]. For pairwise-independent ensembles and arbitrary channels, Cocco *et al.* showed in [15] that the probability that a code in the ensemble has an exponent smaller than a lower bound on the TRC exponent is vanishingly small.

The main motivation of our work is the fact that the aforementioned results highlight the importance of the statistical properties of the error probability and the error exponent across the random-coding ensemble. After describing the main performance metrics of random codes for reliable communication in the next section, namely the error probability and the error exponent, we use the notion of convergence in probability and convergence in distribution to obtain a number of concentration results of such performance metrics, seen as sequences of random variables, as the blocklength tends to infinity. Specifically:

- In Theorem 1 we show that the error exponent of a randomly chosen code from the ensemble converges in probability to the TRC exponent.
- In Theorems 2–4 we provide bounds on the rate of such convergence.
- For codes with a constant number of codewords, Theorem 5 shows that the error exponent of a randomly chosen

code from the ensemble converges in distribution to a Gaussian-like distribution. For codes with a growing sub-exponential number of codewords, Theorem 6 shows that the error exponent converges in distribution to a Gaussian instead.

The aforementioned results are stated in Sec. III, and are valid for the DMC and the i.i.d. and constant-composition ensembles. In addition, for general channels we obtain in Sec. IV the following results:

- For any channel and capacity-achieving ensemble, Theorem 7 states that the error probability of a randomly generated code converges in probability to the ensemble average.
- Theorems 8–10 and Corollary 1 discuss several convergence results relating randomness properties of the error probability, the random-coding error exponent and the TRC exponent.
- Sufficient conditions for the union bound on the error probability and any general function of the error probability to converge to a Gaussian are respectively described in Theorems 11 and 12.

Throughout the paper we use the following notation. Given two positive sequences $\{a_n\}$ and $\{b_n\}$, $n \in \mathbb{N}$, $a_n \doteq b_n$ indicates that $\lim_{n\to\infty} \frac{1}{n}\log\left(\frac{a_n}{b_n}\right) = 0$, $a_n \dot{\leq} b_n$ indicates that $\lim_{n\to\infty} \frac{1}{n}\log\left(\frac{a_n}{b_n}\right) \leq 0$, and the expression $a_n \dot{\geq} b_n$ is similarly defined. The relative entropy between distributions P and Q is denoted as $D(P\|Q)$. A sequence of random variables $\{A_n\}_{n=1}^{\infty}$ converges to A in probability, denoted as $A_n \xrightarrow{\text{(p)}} A$ if for all $\delta > 0$ [16, Sec. 2.2],

$$\lim_{n\to\infty} \mathbb{P}[|A_n - A| > \delta] = 0. \tag{1}$$

If $A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, where $X_i, i = 1, \ldots, n$ are i.i.d. random variables, then $A = \mathbb{E}[X_1]$ and (1) reduces to the weak law of large numbers [16, Th. 2.2.3]. We say that a sequence of random variables $\{A_n\}_{n=1}^{\infty}$ converges to A in distribution, denoted as $A_n \xrightarrow{\text{(d)}} A$ if [16, Sec. 3.2], for all continuity points x of $\mathbb{P}[A \leq x]$, it holds that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \big|\mathbb{P}[A_n \leq x] - \mathbb{P}[A \leq x]\big| = 0. \tag{2}$$

Finally, we say that a sequence $\{A_n\}_{n=1}^{\infty}$ converges almost surely to a constant value A if

$$\mathbb{P}\left[\lim_{n\to\infty} A_n = A\right] = 1, \tag{3}$$

implying that the events for which A_n does not converge to 0 have asymptotically no probability.

This paper is structured as follows. We state our main results for i.i.d. and constant-composition ensembles over DMCs in Sec. III. Additional results for general channels are stated in Sec. IV. The proofs of our theorems are included in Sec. V, while most lemmas thereby used are proved in the Appendix.

II. Preliminaries

We consider the problem of transmitting M_n equiprobable messages over a DMC with transition probability W and finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. We employ a codebook $c_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{M_n}\}$ with $\boldsymbol{x}_m \in \mathcal{X}^n$, for $m = 1, \ldots, M_n$. The channel transforms the transmitted codeword $\boldsymbol{x} \in c_n$ into a channel output $\boldsymbol{y} \in \mathcal{Y}^n$ according to the random transmformation $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} W(y_i|x_i)$. We consider maximum-likelihood decoding, that is, we estimate the transmitted codeword as $\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\in c_n} W^n(\boldsymbol{y}|\boldsymbol{x})$. The error probability is

$$P_{\mathrm{e}}(c_n) = \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{P}\Bigg[\bigcup_{\bar{m}\neq m} \{\boldsymbol{x}_m \to \boldsymbol{x}_{\bar{m}}\}\Bigg], \tag{4}$$

where $\{\boldsymbol{x}_m \to \boldsymbol{x}_{\bar{m}}\} = \{\boldsymbol{y} \in \mathcal{Y} : W^n(\boldsymbol{y}|\boldsymbol{x}_{\bar{m}}) \geq W^n(\boldsymbol{y}|\boldsymbol{x}_m)\}$ is the pairwise error event, i.e., the event of deciding in favor of codeword $\boldsymbol{x}_{\bar{m}}$ when codeword $\boldsymbol{x}_m$ was transmitted. The error exponent of code c_n is defined as

$$E_n(c_n) = -\frac{1}{n}\log P_{\mathrm{e}}(c_n). \tag{5}$$

Let $R = \liminf_{n\to\infty} \frac{1}{n}\log M_n$ be the rate of the code in bits per channel use. An error exponent $E(R)$ is said to be achievable when there exists a sequence of codes $\{c_n\}_{n=1}^{\infty}$ such that $\liminf_{n\to\infty} E_n(c_n) \geq E(R)$. The channel capacity C is the supremum of the code rates R such that $E(R) > 0$.

We next consider the random generation of the codebook. Similarly to random variables, $\mathcal{C}_n$ denotes a random code, and c_n denotes a specific code in the ensemble. In particular, we consider the pairwise-independent random-coding ensemble, i.e., the set of random codes $\mathcal{C}_n$ whose codewords $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{M_n}$ are pairwise-independently generated. We consider the i.i.d. ensemble, in which each codeword is generated according to the distribution

$$Q_{\mathrm{iid}}^n(\boldsymbol{x}) = \prod_{i=1}^{n} Q(x_i), \tag{6}$$

with Q being the input distribution of each symbol, and the constant-composition ensemble, in which each codeword is generated according to the distribution

$$Q_{\mathrm{cc}}^n(\boldsymbol{x}) = \frac{1}{|\mathcal{T}_n(Q_n)|}\mathbb{1}\{\boldsymbol{x} \in \mathcal{T}_n(Q_n)\}, \tag{7}$$

where $\mathcal{T}_n(Q_n)$ is the type class of composition $Q_n \in \mathcal{P}_n(\mathcal{X})$, i.e., all n-length sequences whose empirical distribution is Q_n such that $\max_x |Q_n(x) - Q(x)| \leq \frac{1}{n}$ for a given distribution Q. For a given input distribution or composition Q, we define the random-coding error exponent $E_{\mathrm{rce}}(R, Q)$ as

$$E_{\mathrm{rce}}(R, Q) = \lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n, Q)], \tag{8}$$

where $P_{\mathrm{e}}(\mathcal{C}_n, Q)$ denotes the error probability of the random code ensemble $\mathcal{C}_n$ parametrized by the distribution or composition Q and where the expectation is taken over the code ensemble. The existence of the limit in (8) is known by [17]. Eq. (8) suggests that $E_{\mathrm{rce}}(R, Q)$ is the asymptotic exponent of the ensemble-average probability of error. For i.i.d. code ensembles, Gallager [3, Th. 1] provided an expression of $E_{\mathrm{rce}}(R, Q)$. For constant composition ensembles, the expression of $E_{\mathrm{rce}}(R, Q)$ is provided in [17], [18]. It is known that for any given Q, $E_{\mathrm{rce}}^{\mathrm{iid}}(R, Q) \leq E_{\mathrm{rce}}^{\mathrm{cc}}(R, Q)$ (see e.g., [19]);

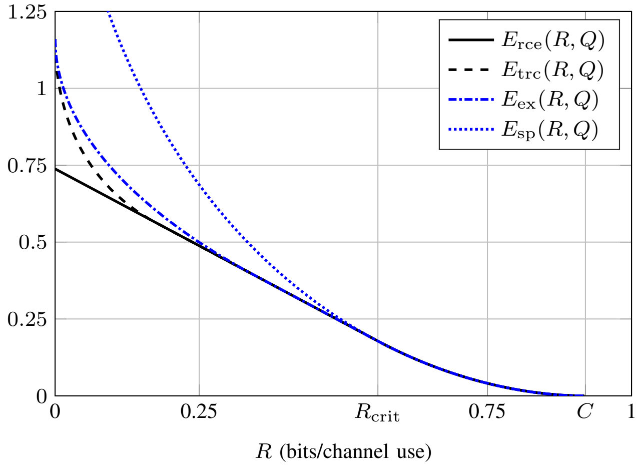


Fig. 1. Example of the rancom-coding error exponent $E_{\rm rce}(R,Q)$ in (8), the typical error exponent $E_{\rm trc}(R,Q)$ in (9), the expurgated error exponent $E_{\rm ex}(R,Q)$ in [3, Eq. (5.7.11)] and the sphere-packing exponent $E_{\rm sp}(R,Q)$ in [22, Eq. (5.8.2)] all for i.i.d. codes over the BSC with equiprobable input and crossover probability 0.01. For this channel, the capacity is $C = 0.919207$ and the critical rate is $R_{\rm crit} = 0.559122$.

when both exponents are optimized over the distribution or composition Q, they coincide.

While $E_{\rm rce}(R,Q)$ in (8) is the limiting exponential rate of decay of the expected probability of error, the typical random-coding exponent $E_{\rm trc}(R,Q)$ is instead defined as the limiting expected error exponent, that is,

$$E_{\rm trc}(R,Q) = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}(\mathcal{C}_n,Q)\big]. \tag{9}$$

Jensen's inequality implies that the random-coding error exponent in (8) and the typical random-coding error exponent in (9) satisfy $E_{\rm rce}(R,Q) \leq E_{\rm trc}(R,Q)$. For i.i.d. codes over the BSC [7], the typical error exponent has been expressed as

$$E_{\rm trc}(R,Q) = \max\{E_{\rm ex}(2R,Q) + R, E_{\rm rce}(R,Q)\}, \tag{10}$$

where $E_{\rm ex}(R,Q)$ is Gallager's expurgated error exponent for i.i.d. ensembles [3, Eq. (5.7.11)]. Since Gallager's expurgated exponent can be smaller than the random coding exponent, we will assume in this paper that for i.i.d. ensembles, whenever we refer to the expurgated exponent we mean $\max\{E_{\rm ex}(R,Q), E_{\rm rce}(R,Q)\}$. For the constant composition ensemble and the general DMC channel, the expurgated exponent derived in [20] (see also [9], [21]) does not exhibit this limitation and the corresponding expression is [9, Eq. (3)]

$$E_{\rm trc}(R,Q) = E_{\rm ex}(2R,Q) + R. \tag{11}$$

We define $R_{\rm crit}$ as the critical rate, the smallest R such that the random coding exponent $E_{\rm rce}(R,Q)$ is tight, i.e., it coincides with the upper bound given by the sphere-packing exponent $E_{\rm sp}(R,Q)$ given in [22, Eq. (5.8.2)]. We show in Figure 1 an example of the aforementioned error exponents for the BSC.

III. Discrete Memoryless Channels

In this section, we introduce our main concentration results for DMCs. Our first result states the convergence in probability of the sequence of error exponents $\{E_n(\mathcal{C}_n)\}_{n=1}^\infty$ to the TRC exponent $E_{\rm trc}(R)$. Since the exponent of the probability of error is not a sum of i.i.d. terms, the weak law of large numbers cannot be applied. This result holds for i.i.d. and constant-composition ensembles over DMCs with input distribution or composition Q.

Theorem 1: For a DMC channel, i.i.d. and constant-composition ensembles with input distribution or composition Q, it holds that

$$E_n(\mathcal{C}_n) \xrightarrow{\rm (p)} E_{\rm trc}(R,Q). \tag{12}$$

for all rates $R \in [R_{\rm crit}, C)$ such that $E_{\rm trc}(R,Q) = E_{\rm rce}(R,Q)$.

Proof: Sec. V-A. ■

Theorem 1 shows the convergence of the sequence of random variables $\{E_n(\mathcal{C}_n)\}_{n=1}^\infty$ to its statistical mean, the TRC exponent. In proving convergence, Theorem 1 shows the achievability of the TRC exponent as well as the fact 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 behaviour of the ensemble than the traditional derivation of the random coding error exponent, which only computes the exponent of the ensemble average error probability. This way, the TRC emerges as the error exponent for i.i.d. and constant-composition ensembles over DMC channels. At zero rate, the TRC and the expurgated exponent coincide for both ensembles. At low, but positive rates (i.e., $0 < R < R_{\rm crit}$), the TRC exponent is lower than or equal to the expurgated exponent and can in some case be strictly smaller. This implies that the codes in the pairwise independent ensemble that achieve the expurgated exponent are not typical codes and are unlikely to be found by random generation.

A refined analysis to that of Theorem 1 consists of studying, separately, the probability tails involved in the definition of convergence in probability in (1). The work in [12], addressed this issue for the constant-composition ensemble over DMCs. Specifically, [12, Theorems 1,2] showed an interesting asymmetry: the probability $\mathbb{P}[E_n(\mathcal{C}_n) < E_{\rm trc}(R,Q) - \varepsilon]$ decays exponentially, while $\mathbb{P}[E_n(\mathcal{C}_n) > E_{\rm trc}(R,Q) + \varepsilon]$ decays double-exponentially. The exponential and double-exponential decay behaviors can be explained by Sanov's theorem in large deviation theory. For our problem, $P_{\rm e}(\mathcal{C}_n)$ —but not $E_n(\mathcal{C}_n)$— is a sum of pairwise random variables, which explains the asymmetric behaviors of the two tails. This result implies that, beyond the concentration property, it is significantly more difficult to find a code in the ensemble with exponent higher than $E_{\rm trc}(R,Q)$.

We next derive some results on the convergence rate of the error exponent $E_n(\mathcal{C}_n)$ to the typical random-coding exponent $E_{\rm trc}(R,Q)$.

Theorem 2: For the i.i.d. ensemble with rate $0 \leq R < C$ and any $\varepsilon > 0$, it holds that

$$\mathbb{P}\big[E_n(\mathcal{C}_n) < E_{\rm trc}(R,Q) - \varepsilon\big] \ \dot{\leq}\ 2^{-n\varepsilon}, \tag{13}$$

$$\mathbb{P}\big[E_n(\mathcal{C}_n) > E_{\rm trc}(R,Q) + \varepsilon\big] = O\bigg(\frac{1}{\sqrt{n}}\bigg). \tag{14}$$

Proof: Sec. V-B. ■

In contrast to the work in [12], the error probability $P_{\rm e}(c_n) = 2^{-nE_n(\mathcal{C}_n)}$ is not a sum of pairwise-independent random variables but a sum of dependent random variables. Refining the bound in (14) to obtain a double exponential decay as in the constant-compostion case remains a challenging problem.

Theorem 2 strengthens Theorem 1. The Berry-Esseen theorem [16, Theorem 3.4.17] and Theorem 6 are used to obtain (14). For a fixed code ensemble c_n, we define the union bound to the error probability as,

$$P_{\rm e}^{\rm ub}(c_n) = \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j], \tag{15}$$

whose exponent is

$$E_n^{\rm ub}(c_n) = -\frac{1}{n}\log P_{\rm e}^{\rm ub}(c_n). \tag{16}$$

For the above union bound and low rates, we refine the upper tail bound in (14) as follows.

Theorem 3: For all rates R such that $E_{\rm trc}(R,Q) > E_{\rm rce}(R,Q)$, for any $\epsilon > 0$ and large enough n the sequence of random variables $\{E_n^{\rm ub}(\mathcal{C}_n)\}_{n=1}^{\infty}$ satisfies:

$$\mathbb{P}\left[E_n^{\rm ub}(\mathcal{C}_n) \geq E_{\rm ex}(R,Q) + \epsilon\right] \leq 2^{-2^{n\epsilon}}. \tag{17}$$

Proof: Sec. V-C. ■

Theorem 3 states that the probability to find a code in the i.i.d. ensemble for which $E_n^{\rm ub}$ is larger than the expurgated exponent tends to zero double-exponentially fast in n. In [12] it is shown that, for constant composition codes, the probability to find a code whose exponent E_n exceeds the expurgated exponent decays double-exponentially. This fact together with Theorem 3 suggest, although not proven here, that it is highly unlikely to find a code in the i.i.d. ensemble whose exponent exceeds the expurgated exponent. In Theorem 4 below we show that this is indeed the case at least for $R = 0$. The proof of Theorem 3 is similar in spirit to [12, Th. 2], the differences being detailed in Sec. V-C and Appendix A.10.

Theorem 4: For the i.i.d. or constant-composition ensembles with rate $R = 0$ and any $\varepsilon > 0$, we have that

$$\mathbb{P}\left[E_n(\mathcal{C}_n) \geq E_{\rm trc}(0,Q) + \epsilon\right] \leq 2^{-2^{n\epsilon}}. \tag{18}$$

Proof: Sec. V-D. ■

Theorem 4 shows that, at least for the point $R = 0$, the probability of finding a code from the i.i.d. or constant-composition ensembles with an exponent larger than $E_{\rm trc}(0,Q) = E_{\rm ex}(0,Q)$ decays double-exponentially in n.

So far, we have introduced results related to the convergence in probability of the error exponent for i.i.d. and constant composition ensembles. In the remainder of the section, we discuss the convergence in distribution of the sequence of error exponent random variables $\{E_n(\mathcal{C}_n)\}_{n=1}^{\infty}$ as $n \to \infty$ for vanishingly small rates. Theorem 5 and Theorem 6 below are valid for i.i.d. codes and for constant-composition codes as long as the type Q_X satisfies

$$|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|\left(1 - \frac{|\mathcal{T}_n(Q_{XX'}^*)|}{|\mathcal{T}_n(Q_X)|^2}\right) \to 0 \tag{19}$$

as $n \to \infty$, where $Q_{XX'}^* = Q_X Q_X$. We will call types Q_X fulfilling (19) *regular* types.

Theorem 5: Let $M_n = M$ be a constant number of messages, fixed for every n, and let $U_{ij} \sim \mathcal{N}(0,1)$, for $i = 1,\dots,M$ and $j = 1,\dots,M$ such that $i \neq j$, be a set of independent standard normal random variables. For i.i.d. codes or constant-composition codes with the type Q_X satisfying (19), the error exponent sequence $\{E_n(\mathcal{C}_n)\}_{n=1}^{\infty}$ for both i.i.d. and constant-composition random-coding ensembles with regular type satisfies

$$\frac{E_n(\mathcal{C}_n) - \mathbb{E}[E_n(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(E_n(\mathcal{C}_n))}} \xrightarrow{\rm (d)} \frac{\min_{i\neq j} U_{ij} - \mathbb{E}[\min_{i\neq j} U_{ij}]}{\sqrt{\mathrm{Var}(\min_{i\neq j} U_{ij})}}. \tag{20}$$

Proof: Sec. V-E. ■

We illustrate in Fig. 2 the histogram of the error exponent $E_n(\mathcal{C}_n)$ used over a binary symmetric channel (BSC) with crossover probability $p = 0.11$, equiprobable input and $M = 4$ codewords for a blocklength of $n = 10,000$. The histograms are obtained for the i.i.d. and constant-composition ensembles using the Monte Carlo method with 10^7 trials. For the sake of comparison, we also depict the asymptotic distribution of the random variable $\min_{i\neq j} U_{ij}$ in the right-hand side of (20) (solid), and a normal approximation with the same mean and variance (dashed). We observe that the two histograms match the asymptotic distribution on the right-hand side of (20). When comparing with the Gaussian approximation, is a noticeable difference in the two tails. We refer to the distribution on the right-hand side of (20) as Gaussian-like.

Theorem 5, valid for an exactly constant number of messages, states that the random-coding error exponent converges to a Gaussian-like distribution. In Theorem 6 below we let the number of messages M_n grow sub-exponentially with n, and show that as long as $M_n \gg \sqrt{n}$, the error exponent sequence $\{E_n(\mathcal{C}_n)\}_{n=1}^{\infty}$ converges to a Gaussian.

Theorem 6: Let M_n be a subexponential number of messages, namely $\lim_{n\to\infty}\frac{1}{n}\log M_n = 0$, satisfying

$$\sum_{n=1}^{\infty}\frac{1}{M_n(M_n-1)} < \infty. \tag{21}$$

Then, the the error exponent sequence $\{E_n(\mathcal{C}_n)\}_{n=1}^{\infty}$ for i.i.d. and constant-composition ensembles satisfies

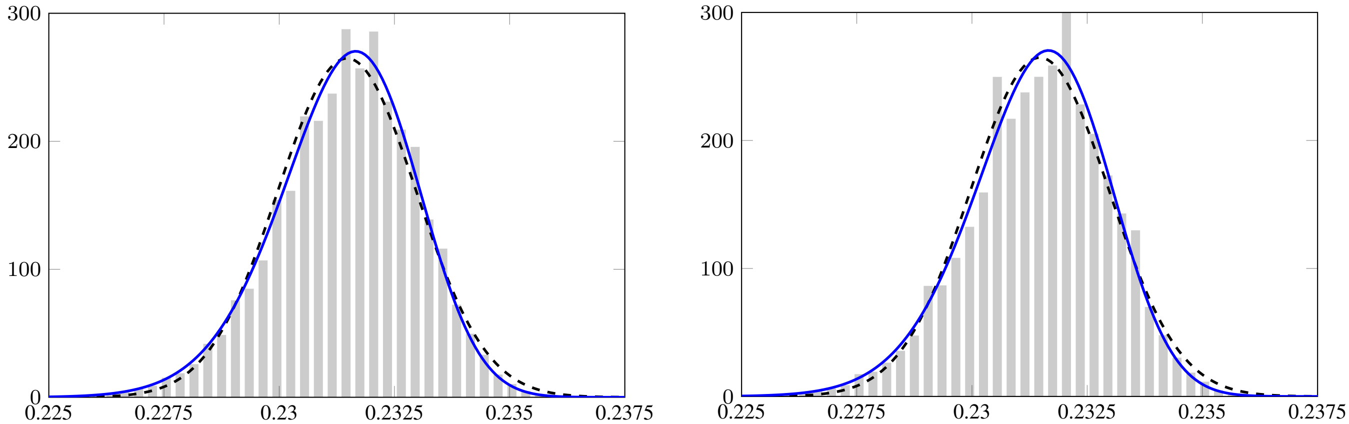
$$\frac{E_n(\mathcal{C}_n) - \mathbb{E}[E_n(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(E_n(\mathcal{C}_n))}} \xrightarrow{\rm (d)} \mathcal{N}(0,1). \tag{22}$$

Proof: Sec. V-F. ■

For a constant number of messages $M_n = M$, the condition (21) in Theorem 6 is not satisfied, and therefore the error exponent sequence does not concentrate according to (22) but to (20) instead. For example, when the number of messages is such that $M_n = \Omega\big(n^{\frac{1+\delta}{2}}\big)$, the condition (21) is satisfied and therefore the error exponent sequence converges to (22).

IV. General Channels

In this section, we introduce a number of new results related to the concentration of the error probability and error exponent for relatively general channels and ensembles.



(a) i.i.d. (c) constant-composition

Fig. 2. 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 $\min_{i\neq j} U_{ij}$.

Specifically, Theorem 7 applies to any channel for which the strong converse holds. The rest of the results in this section hold for any channel, discrete or continuous, with and without memory and pairwise independent ensembles with the following exceptions: Corollary 1 and Theorem 12, hold for ensembles satisfying a certain condition and Theorem 11 which holds only for i.i.d. ensembles. The first result is a direct consequence of elementary probability results such as Chebyshev's inequality or Jensen's inequality.

Theorem 7: For a channel and random-coding ensemble such that $\mathbb{E}[P_{\rm e}(\mathcal{C}_n)] \to 0$ as $n \to \infty$ for $0 \le R < C$ and $\mathbb{E}[P_{\rm e}(\mathcal{C}_n)] \to 1$ for $R > C$, the error probability sequence $\{P_{\rm e}(\mathcal{C}_n)\}_{n=1}^{\infty}$ satisfies

$$P_{\rm e}(\mathcal{C}_n) - \mathbb{E}[P_{\rm e}(\mathcal{C}_n)] \xrightarrow{\rm (p)} 0. \tag{23}$$

Proof: Sec. V-G. ■

The following theorem has implications on the convergence of the error exponent to a Gaussian r.v., as show in Corollary 1:

Theorem 8: For a general channel and a pairwise-independent ensemble, under the condition that

$$\frac{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)^2]}{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]^2} \to 1, \tag{24}$$

we have that the error exponent sequence $\{E_n(\mathcal{C}_n)\}_{n=1}^{\infty}$ satisfies

$$E_n(\mathcal{C}_n) \xrightarrow{\rm (p)} E_{\rm trc}(R). \tag{25}$$

Proof: Sec. V-I. ■

Theorem 8 has the following corollary:

Corollary 1: For a general channel and pairwise-independent ensemble such that the normalized error probability converges in distribution to the standard normal distribution, that is

$$\frac{P_{\rm e}(\mathcal{C}_n) - \mathbb{E}[P_{\rm e}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\rm e}(\mathcal{C}_n))}} \xrightarrow{\rm (d)} \mathcal{N}(0,1), \tag{26}$$

we have that

$$E_n(\mathcal{C}_n) \xrightarrow{\rm (p)} E_{\rm trc}(R). \tag{27}$$

Proof: Sec. V-K. ■

We remark that condition (26) is sufficient, but not necessary for the convergence in probability.

The next result is based on [15, Th. 1] and the Paley-Zygmund inequality [23, p. 1] and has implications on the convergence of the exponent to a Gaussian r.v., as shown by Theorem 10:

Theorem 9: For a general channel and pariwise-independent ensemble with rate such that $E_{\rm trc}(R) > E_{\rm rce}(R)$, we have

$$\frac{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]^2}{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)^2]} \to 0. \tag{28}$$

Proof: Sec. V-H. ■

For low rates, where the typical random-coding error exponent is strictly larger than the random-coding error exponent, the second-order moment of the error probability vanishes slower than the squared first-order moment. This implies that $\mathrm{Var}\left(P_{\rm e}(\mathcal{C}_n)\right)$ vanishes slower than the squared ensemble average $\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]^2$, suggesting that the error probability cannot converge to a Gaussian distribution in this rate regime. Such intuition is formalized in the next results, based on Theorem 9 and Slutsky's theorem [24, p. 334].

Theorem 10: For any code ensemble and channel such that $E_{\rm trc}(R) > E_{\rm rce}(R)$, it holds that the error probability sequence $\{P_{\rm e}(\mathcal{C}_n)\}_{n=1}^{\infty}$ satisfies

$$\frac{P_{\rm e}(\mathcal{C}_n) - \mathbb{E}[P_{\rm e}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\rm e}(\mathcal{C}_n))}} \overset{\rm (d)}{\not\longrightarrow} \mathcal{N}(0,1). \tag{29}$$

Proof: Sec. V-J. ■

In the remainder of the section, we state two auxiliary results related to the convergence in distribution of the union bound to the error probability of a code c_n in (15), and the convergence in distribution of an arbitrary function of the error probability.

Theorem 11: Let Y_{12} and γ^2 be two parameters respectively given by $Y_{12} = \mathbb{P}\big[\{\boldsymbol{X}_1 \to \boldsymbol{X}_2\}\big] - \mathbb{E}\big[\mathbb{P}[\{\boldsymbol{X}_1 \to \boldsymbol{X}_2\}]\big]$ and $\gamma^2 = \mathrm{Var}\big(\mathbb{P}[\{\boldsymbol{X}_1 \to \boldsymbol{X}_2\}]\big)$. For general channels and i.i.d. ensembles such that

$$\frac{M_n}{\gamma^3}\mathbb{E}[|Y_{12}|^3] \to 0\ , \quad \frac{M_n}{\gamma^4}\mathbb{E}[|Y_{12}|^4] \to 0 \tag{30}$$

as $n \to \infty$, we have that the error probability sequence $\{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)\}_{n=1}^{\infty}$ satisfies

$$\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n))}} \xrightarrow{(\mathrm{d})} \mathcal{N}(0,1). \tag{31}$$

Proof: Sec. V-L. ■

While Theorem 11 gives sufficient conditions for the convergence in probability of an upper bound on the error probability, Theorem 10 gives a sufficient condition that prevents this to happen. These results imply that for all codes and channels such that the two conditions (30) hold, the condition $E_{\mathrm{trc}}(R,Q) > E_{\mathrm{rce}}(R,Q)$ cannot be satisfied.

In the last result, we develop a general condition for the convergence in distribution of a random variable sequence to the standard normal random variable.We have been unable to specify for which specific channels and (random) codebook ensembles these conditions hold.

Theorem 12: Let $g_n : [0,1] \to \mathbb{R}$ be an arbitrary sequence of functions. For channels and random codebook ensembles satisfying

$$\mathbb{E}\left[\left|\frac{g_n(P_{\mathrm{e}}(\mathcal{C}_n)) - \mathbb{E}[g_n(P_{\mathrm{e}}(\mathcal{C}_n))]}{\sqrt{\mathrm{Var}(g_n(P_{\mathrm{e}}(\mathcal{C}_n)))}}\right|\right] \to 0, \tag{32}$$

and

$$\mathbb{E}\left[\left|\left(\frac{g_n(P_{\mathrm{e}}(\mathcal{C}_n)) - \mathbb{E}[g_n(P_{\mathrm{e}}(\mathcal{C}_n))]}{\sqrt{\mathrm{Var}(g_n(P_{\mathrm{e}}(\mathcal{C}_n)))}}\right)^2 - 1\right|\right] \to 0, \tag{33}$$

the sequence $\{g_n(P_{\mathrm{e}}(\mathcal{C}_n))\}_{n=1}^{\infty}$ satisfies

$$\frac{g_n(P_{\mathrm{e}}(\mathcal{C}_n)) - \mathbb{E}[g_n(P_{\mathrm{e}}(\mathcal{C}_n))]}{\sqrt{\mathrm{Var}(g_n(P_{\mathrm{e}}(\mathcal{C}_n)))}} \xrightarrow{(\mathrm{d})} \mathcal{N}(0,1). \tag{34}$$

Proof: Sec. V-M. ■

V. Proofs of Theorems

Before proving our main results, we introduce some definitions related to the Stein's method [25] used throughout the proofs. We also propose a novel, modified Wasserstein metric that is used throughout the section. Let $\mathcal{V}$ be the set of positive-valued piece-wise functions $h(u)$ given, for some $c \geq 0$ and $a \in \mathbb{R}$, by

$$h(u) = \begin{cases} c & u \leq a \\ a + c - u & a < u < a + c \\ 0 & u \geq a + c. \end{cases} \tag{35}$$

We next define two probability metrics.

Definition 1: For two random variables X and Y, the probability metrics have the following form:

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

$$\bar{d}_{\mathcal{H}}(X,Y) = \sup_{h\in\mathcal{H}} \min\Big\{\big|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]\big|, \big|\mathbb{E}[h(-X)] - \mathbb{E}[h(Y)]\big|\Big\}, \tag{37}$$

where $\mathcal{H}$ is some family of "test" functions on $\mathbb{R}$.

By taking $\mathcal{H} = \{\mathbb{1}\{\cdot \leq u\}\ u \in \mathbb{R}\}$ in (36) and the probability metric $d_{\mathcal{H}}(X,Y)$, we obtain the Kolmogorov metric, which denote by d_K [25]. By definition, the convergence in the Kolmogorov metric means the convergence in distribution. By taking $\mathcal{H} = \{h : \mathbb{R} \to \mathbb{R} : |h(u) - h(v)| \leq |u - v|\}$ and the probability metric $d_{\mathcal{H}}(X,Y)$, we obtain the Wasserstein metric, which we denote d_W [25]. By taking $\mathcal{H} = \{h \in \mathcal{V} : c \leq 4\sqrt{2\pi}\}$ and the probability metric $d_{\mathcal{H}}(X,Y)$, we obtain a slightly modified Wasserstein metric $\tilde{d}_{W,\mathrm{mod}}$. Finally, by taking $\mathcal{H} = \{h \in \mathcal{V} : c \leq 4\sqrt{2\pi}\}$ and the probability metric $\bar{d}_{\mathcal{H}}(X,Y)$, we obtain a modified Wasserstein metric[1], which we denote $d_{W,\mathrm{mod}}$.

The following auxiliary lemmas whose proof can be found in the Appendix A.1, are key in deriving the convergence in distribution results of this paper.

Lemma 1: Let $\{U_n\}_{n=1}^{\infty}$ be a sequence of random variables such that $U_n \xrightarrow{(\mathrm{d})} U$ for some random variable U. Then, under the condition that $\mathbb{E}[|U_n|^{2+\varepsilon}] < L$ for some $\varepsilon > 0$ and constant $L < \infty$, it holds that

$$\frac{U_n - \mathbb{E}[U_n]}{\sqrt{\mathrm{Var}(U_n)}} \xrightarrow{(\mathrm{d})} \frac{U - \mathbb{E}[U]}{\sqrt{\mathrm{Var}(U)}}. \tag{38}$$

Proof: Appendix A.1. ■

Lemma 2 (De Caen [26]): Let $\{A_i\}_{i\in\mathcal{I}}$ be a finite family of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then[2]

$$\mathbb{P}\bigg(\bigcup_{i\in\mathcal{I}} A_i\bigg) \geq \sum_{i\in\mathcal{I}} \frac{\mathbb{P}^2(A_i)}{\sum_{j\in\mathcal{I}} \mathbb{P}[A_i \cap A_j]}. \tag{39}$$

A. Proof of Theorem 1

The proof of Theorem 1 is structured as follows. From (40) to (42), we prove our theorem for the case that $E_{\mathrm{trc}}(R,Q) = E_{\mathrm{rce}}(R,Q)$, while the rest of the proof is for $E_{\mathrm{trc}}(R,Q) > E_{\mathrm{rce}}(R,Q)$. After introducing some definitions and auxiliary lemmas from (43) to (57), the proof individually studies the three terms of (58): the first term from (59) to (91), the second term from (92) to (112) and the third term from (113) to (115). We end the proof by adapting some of the steps to the constant-composition ensemble in (116).

Lemma 3: Suppose that for the channel considered, $E_{\mathrm{trc}}(R,Q) = E_{\mathrm{rce}}(R,Q)$. Then for the i.i.d. and constant-composition ensembles, it holds that $\lim_{n\to\infty} \mathbb{E}\big[P_{\mathrm{e}}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big] = 2^{-\lambda E_{\mathrm{trc}}(R,Q)}$ for any $\lambda > 0$.

Proof: Appendix A.2. ■

For the case $E_{\mathrm{trc}}(R,Q) = E_{\mathrm{rce}}(R,Q)$, we let $\varphi(\lambda) = 2^{-\lambda E_{\mathrm{trc}}(R,Q)}$ for all $\lambda > 0$ be the Laplace transform of the constant random variable $-E_{\mathrm{trc}}(R,Q)$, and let $\varphi_n(\lambda)$ be the Laplace transform of the distribution of $\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)$, that is, $\varphi_n(\lambda) = \mathbb{E}[2^{\lambda\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)}] = \mathbb{E}\big[P_{\mathrm{e}}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big]$. Then, by Lemma 3, it holds that

$$\lim_{n\to\infty} \varphi_n(\lambda) = 2^{-\lambda E_{\mathrm{trc}}(R,Q)}. \tag{40}$$

Applying the Levy's continuity theorem [27, Sec. XIII.1], we obtain from (40) that

$$-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \xrightarrow{(\mathrm{d})} E_{\mathrm{trc}}(R,Q). \tag{41}$$

[1]This definition of Wasserstein metric is a variant of the definition in [25], where we constraint the set $\mathcal{H}$ to achieve a tighter bound.

[2]We make the convention $\frac{0}{0} = 0$, so that events of probability zero are not counted in (39).

However, we know that the convergence in distribution to a constant implies convergence in probability, i. e.

$$-\frac{1}{n}\log P_{\rm e}(\mathcal{C}_n) \xrightarrow{(\rm p)} E_{\rm trc}(R,Q). \quad (42)$$

We now switch to the range of rates for which $E_{\rm trc}(R,Q) > E_{\rm rce}(R,Q)$. We first need some definitions and lemmas. For this range of rates, the proof uses the union bound to the error probability (4) and shows that it gives a good estimate of the probability of error. The union bound is given by,

$$P_{\rm e}(\mathcal{c}_n) \leq P_{\rm e}^{\rm ub}(\mathcal{c}_n), \quad (43)$$

where $P_{\rm e}^{\rm ub}(\mathcal{c}_n)$ is defined in (15), and we define its finite-length error exponent as

$$E_n^{\rm ub}(\mathcal{c}_n) = -\frac{1}{n}\log P_{\rm e}^{\rm ub}(\mathcal{c}_n). \quad (44)$$

We denote by $E_{\rm trc}(R,Q)$ and $E_{\rm rce}(R,Q)$ respectively the typical error and the random coding error exponents for the fixed underlying distribution Q, and we define

$$d_{\rm B}(x,x') = -\log\left(\sum_y \sqrt{W(y|x)W(y|x')}\right) \quad (45)$$

to be the Bhattacharyya distance between symbols $x, x' \in \mathcal{X}$.

We assume that the DMC is such that

$$0 < D_{\rm b} = \max_{x,x'} d_{\rm B}(x,x') < \infty, \quad (46)$$

that is, we leave the cases where $W(y|x)W(y|x') = 0$ for for some x and x' and all y beyond the scope of the paper. This case would correspond to a positive zero-error capacity, where some symbols cannot be confused at the decoder.

We let $\mathcal{P}_n(\mathcal{X}\times\mathcal{X})$ be the set of all joint types on $\mathcal{X}\times\mathcal{X}$, and $\mathcal{P}(\mathcal{X}\times\mathcal{X})$ be the set of all possible probability distributions on $\mathcal{X}\times\mathcal{X}$. For each $P_{XX'} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X})$, let $\mathcal{N}(P_{XX'})$ be the number of codeword pairs in a specific code $\mathcal{c}_n$ such that their joint type is $P_{XX'}$. Let $Q_X = Q'_X = Q$. Define

$$\mathcal{V}_n = \Big\{\mathcal{N}(P_{XX'}) = 0, \quad \forall P_{XX'} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}) : D(P_{XX'}\|Q_X Q'_X) > 2R\Big\} \quad (47)$$

which is the event that the (random) number of pairs $(i,j) \in [M_n]\times[M_n]$ such that $i \neq j$ and $(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})$ is equal to zero for each n-joint type $P_{XX'}$ with $D(P_{XX'}\|Q_X Q'_X) > 2R$. In addition, define $\bar{E}_{\rm trc}(\nu,R,Q)$ as (49) at the top of next page, where $P^*_{XX'}$ is an optimizer of $\min_{P_{XX'}\in\mathcal{P}} D(P_{XX'}\|Q_X Q_{X'}) + \sum_{x,x'} d_{\rm B}(x,x')P_{XX'}(x,x') - R$.

First, we introduce some auxiliary results about the exponential decay of the pairwise error probability between two codewords, using the method of types.

Lemma 4: For $R < R_{\rm crit}$, the pairwise codeword error probability between two codewords $\boldsymbol{x}_i, \boldsymbol{x}_j$ with joint type $P_{XX'}$ satisfies $\mathbb{P}[\boldsymbol{x}_i \to \boldsymbol{x}_j | P_{XX'}] = g_n(P_{XX'}) := 2^{-n\sum_{x,x'} d_{\rm B}(x,x')P_{XX'}(x,x')}$.

Proof: Appendix A.3. ■

Lemma 5: Recall the definition of $\mathcal{V}_n$ in (47). Let $\mathcal{C}_n$ be a given i.i.d. random codebook ensemble. Then, we have that $\mathbb{P}\big[\mathcal{V}_n^c\big] \dot{\leq} 2^{-n\alpha(R)}$ for some $\alpha(R) > 0$ for all $R \geq 0$.

Proof: Appendix A.4. ■

Lemma 6: Recall the definition of $\bar{E}_{\rm trc}(\nu,R)$ in (49). Assume that $0 < R < R_{\rm crit}$. Take an arbitrary $\nu \geq 0$ such that $\nu \leq 2R$. Let $Q_X = Q_{X'} = Q$ and define $\mathcal{P} = \Big\{P_{XX'} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}) : D(P_{XX'}\|Q_X Q'_X) \leq 2R - \nu\Big\}$. Also let D_n be

$$D_n = \frac{1}{M}\sum_{P_{XX'}\in\mathcal{P}} \mathcal{N}(P_{XX'})g_n(P_{XX'}) \quad (50)$$

where the function $g_n : \mathcal{X}\times\mathcal{X} \to \mathbb{R}_+$ is defined in Lemma 4. Then,

$$\mathbb{E}[D_n] \doteq 2^{-n\bar{E}_{\rm trc}(\nu,R)} \quad (51)$$

and also, we have

$$\frac{{\rm Var}(D_n)}{\big(\mathbb{E}[D_n]\big)^2} \dot{\leq} 2^{-n\nu}. \quad (52)$$

Proof: Appendix A.5. ■

Lemma 7: Let

$$E_{\rm trc}^{\rm ub}(R,Q) = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)]. \quad (53)$$

Then, for $0 < R < R_{\rm crit}$, the following holds:

$$\begin{aligned} &E_{\rm trc}^{\rm ub}(R,Q) \\ &= \min_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X}):D(P_{XX'}\|Q_X Q'_X)\leq 2R} D(P_{XX'}\|Q_X Q'_X) \\ &\quad + \sum_{x,x'} d_{\rm B}(x,x')P_{XX'}(x,x') - R \quad (54) \\ &= \bar{E}_{\rm trc}(0,R,Q), \quad (55) \end{aligned}$$

where $\bar{E}_{\rm trc}$ is defined in (49), Lemma 6.

Proof: Appendix A.6. ■

Lemma 8: Consider the range of rates $0 \leq R < R_{\rm crit}$ such that $E_{\rm trc}^{\rm ub}(R,Q) > E_{\rm rce}(R,Q)$. Then, for any $\varepsilon > 0$, there exists some $\kappa > 0$ such that

$$\mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n) > \frac{1}{2}2^{-n(E_{\rm trc}^{\rm ub}(R,Q)-\varepsilon)}\Big] + \mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n) < 2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\varepsilon)}\Big] \dot{\leq} \frac{1}{n^{1+\kappa}}. \quad (56)$$

Proof: Appendix A.7. ■

Lemma 9: For all rate R such that $0 < R < R_{\rm crit}$ and for some $\delta(R) > 0$, it holds that

$$0 \leq \frac{\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]}{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]} - 1 \leq 2^{-n\big(\delta(R)+E_{\rm trc}^{\rm ub}(R,Q)-E_{\rm rce}(R,Q)\big)}. \quad (57)$$

Proof: Appendix A.8. ■

$$\bar{E}_{\rm trc}(\nu,R,Q) := \min_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X}):D(P_{XX'}\|Q_XQ'_X)\le 2R-\nu} \Big(D(P_{XX'}\|Q_XQ'_X)+\sum_{x,x'} d_{\rm B}(x,x')P_{XX'}(x,x')-R\Big) \tag{48}$$

$$= \begin{cases} R+\sum_{x,x'} d_{\rm B}(x,x')P^*_{XX'}(x,x'), & D(P^*_{XX'}\|Q_XQ'_X)=2R-\nu, \\ E_{\rm rce}(R,Q) & \text{otherwise} \end{cases}, \tag{49}$$

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 $E_{\rm trc}(R,Q)$ can be written and upper bounded as

$$\begin{aligned}
&\mathbb{P}\big[|E_n(\mathcal{C}_n)-E_{\rm trc}(R)|>3\varepsilon\big] \\
&\quad\le \underbrace{\mathbb{P}\Big[\big|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\big|>\varepsilon\Big]}_{\alpha_n} \\
&\quad+\underbrace{\mathbb{P}\bigg[\bigg|E_n^{\rm ub}(\mathcal{C}_n)-\Big(-\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]\Big)\bigg|>\varepsilon\bigg]}_{\beta_n} \\
&\quad+\underbrace{\mathbb{P}\bigg[\bigg|\Big(-\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]\Big)-E_{\rm trc}(R,Q)\bigg|>\varepsilon\bigg]}_{\gamma_n}.
\end{aligned} \tag{58}$$

We next show that the terms α_n, β_n and γ_n in (58) tend to zero as $n\to\infty$, implying the concentration in (12).

1) First term of (58)*:* The term α_n quantifies the deviation of the error exponent of the error probability (5) from the union bound (15). By the symmetry of the pairwise-independent 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

$$\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_i\to\boldsymbol{X}_j]\big]=\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_1\to\boldsymbol{X}_2]\big]. \tag{59}$$

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

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{P}\big[\{\boldsymbol{X}_i\to\boldsymbol{X}_j\}\cap\{\boldsymbol{X}_i\to\boldsymbol{X}_k\}\big]\Big] \\
&\qquad=\mathbb{E}\Big[\mathbb{P}\big[\{\boldsymbol{X}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{X}_1\to\boldsymbol{X}_3\}\big]\Big]
\end{aligned} \tag{60}$$

where in both (59) and (60), 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 $0<R<R_{\rm crit}(Q)$.

For the case of $R=0$, we first observe that the union bound (15) can be bounded from above as

$$P_{\rm e}^{\rm ub}(\mathcal{C}_n)=\frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\mathbb{P}[\boldsymbol{x}_i\to\boldsymbol{x}_j] \tag{61}$$

$$\le (M_n-1)\max_{i\neq j}\mathbb{P}[\boldsymbol{x}_i\to\boldsymbol{x}_j], \tag{62}$$

while the probability of error (4) can be lower bounded by

$$P_{\rm e}(\mathcal{C}_n)\ge\frac{1}{M_n}\max_{i\neq j}\mathbb{P}[\boldsymbol{x}_i\to\boldsymbol{x}_j]. \tag{63}$$

From (62) and (63), we have that the first term in the r.h.s. of (58) satisfies

$$\alpha_n=\mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)>2^{n\varepsilon}P_{\rm e}(\mathcal{C}_n)\Big] \tag{64}$$

$$\le\mathbb{P}\Big[(M_n-1)\max_{i\neq j}\mathbb{P}[\boldsymbol{X}_i\to\boldsymbol{X}_j]> 2^{n\varepsilon}\frac{1}{M_n}\max_{i\neq j}\mathbb{P}[\boldsymbol{X}_i\to\boldsymbol{X}_j]\Big] \tag{65}$$

$$=\mathbb{1}\Big\{(M_n-1)>2^{n\varepsilon}\frac{1}{M_n}\Big\}. \tag{66}$$

Since M_n is any sub-exponential sequence in n, the expression in (66) vanishes as $n\to\infty$ for $\varepsilon>0$.

We now consider the case of $0<R<R_{\rm crit}(Q)$. We define the sequence $a_n\triangleq 2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\frac{\varepsilon}{2})}$. Then, we have

$$\begin{aligned}
&\mathbb{P}\bigg[\bigg|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\bigg|>\varepsilon\bigg] \\
&\quad=\mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)-a_n-2^{\varepsilon n}\big(P_{\rm e}(\mathcal{C}_n)-a_n\big)>(2^{\varepsilon n}-1)a_n\Big]
\end{aligned} \tag{67}$$

$$\begin{aligned}
&\le\mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)-a_n>\frac{1}{2}(2^{\varepsilon n}-1)a_n\Big] \\
&\qquad+\mathbb{P}\Big[-2^{\varepsilon n}\big(P_{\rm e}(\mathcal{C}_n)-a_n\big)>\frac{1}{2}(2^{\varepsilon n}-1)a_n\Big],
\end{aligned} \tag{68}$$

where (68) follows from

$$\begin{aligned}
\mathbb{P}[A+B>2C]&\le\mathbb{P}[\{A>C\}\cup\{B>C\}] \\
&\le\mathbb{P}[A>C]+\mathbb{P}[B>C].
\end{aligned} \tag{69}$$

Now, observe that

$$\mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)-a_n>\frac{1}{2}(2^{\varepsilon n}-1)a_n\Big]$$

$$=\mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)>\frac{1}{2}(2^{\varepsilon n}+1)2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\varepsilon/2)}\Big] \tag{70}$$

$$\le\mathbb{P}\Big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)>\frac{1}{2}2^{-n(E_{\rm trc}^{\rm ub}(R,Q)-\varepsilon/2)}\Big]. \tag{71}$$

On the other hand, from (69) we also have

$$\begin{aligned}
&\mathbb{P}\Big[-2^{\varepsilon n}\big(P_{\rm e}(\mathcal{C}_n)-a_n\big)>\frac{1}{2}(2^{\varepsilon n}-1)a_n\Big] \\
&\le\mathbb{P}\Big[2^{\varepsilon n}(P_{\rm e}^{\rm ub}(\mathcal{C}_n)-P_{\rm e}(\mathcal{C}_n))>\frac{1}{4}(2^{\varepsilon n}-1)a_n\Big] \\
&\qquad+\mathbb{P}\Big[-2^{\varepsilon n}\Big(P_{\rm e}^{\rm ub}(\mathcal{C}_n)-a_n\Big)>\frac{1}{4}(2^{\varepsilon n}-1)a_n\Big].
\end{aligned} \tag{72}$$

Now, we have

$$\mathbb{P}\bigg[-2^{\varepsilon n}\bigg(P_{\rm e}^{\rm ub}(\mathcal{C}_n)-a_n\bigg)>\frac{1}{4}(2^{\varepsilon n}-1)a_n\bigg]$$
$$=\mathbb{P}\bigg[P_{\rm e}^{\rm ub}(\mathcal{C}_n)<\bigg(1-\frac{1}{4}\bigg(\frac{2^{\varepsilon n}-1}{2^{\varepsilon n}}\bigg)\bigg)2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\varepsilon/2)}\bigg] \tag{73}$$
$$\leq\mathbb{P}\bigg[P_{\rm e}^{\rm ub}(\mathcal{C}_n)<2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\varepsilon/2)}\bigg]. \tag{74}$$

In addition, we also have

$$\mathbb{P}\bigg[2^{\varepsilon n}(P_{\rm e}^{\rm ub}(\mathcal{C}_n)-P_{\rm e}(\mathcal{C}_n))>\frac{1}{4}(2^{\varepsilon n}-1)a_n\bigg]$$
$$\dot{\leq}\ a_n^{-1}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)-P_{\rm e}(\mathcal{C}_n)] \tag{75}$$
$$=2^{(E_{\rm trc}^{\rm ub}(R,Q)+\varepsilon/2)n}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)-P_{\rm e}(\mathcal{C}_n)], \tag{76}$$

where (75) follows from $P_{\rm e}^{\rm ub}(\mathcal{C}_n)\geq P_{\rm e}(\mathcal{C}_n)$ and Markov's inequality, and (76) follows from the definition of the sequence a_n.

Now, for $R>0$ and $E_{\rm trc}^{\rm ub}(R,Q)>E_{\rm rce}(R,Q)$, from Lemma 9, we have

$$\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)-P_{\rm e}(\mathcal{C}_n)]$$
$$=\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]\bigg(\frac{\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]}{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]}-1\bigg) \tag{77}$$
$$\dot{\leq}\ 2^{-nE_{\rm rce}(R,Q)}\bigg(2^{-n\big(\delta(R)+E_{\rm trc}^{\rm ub}(R,Q)-E_{\rm rce}(R,Q)\big)}\bigg). \tag{78}$$

From (76) and (78), we obtain

$$\mathbb{P}\bigg[2^{\varepsilon n}(P_{\rm e}^{\rm ub}(\mathcal{C}_n)-P_{\rm e}(\mathcal{C}_n))>\frac{(2^{\varepsilon n}-1)}{4}a_n\bigg]$$
$$\dot{\leq}\ 2^{(E_{\rm trc}^{\rm ub}(R,Q)+\frac{\varepsilon}{2})n}2^{-nE_{\rm rce}(R,Q)}$$
$$\times 2^{-n\big(\delta(R)+E_{\rm trc}^{\rm ub}(R,Q)-E_{\rm rce}(R,Q)\big)} \tag{79}$$
$$\dot{\leq}\ 2^{-n\big(\delta(R)-\varepsilon/2\big)}. \tag{80}$$

Hence, from (72), (74), and (80), we have

$$\mathbb{P}\bigg[-2^{\varepsilon n}\big(P_{\rm e}(\mathcal{C}_n)-a_n\big)>\frac{1}{2}(2^{\varepsilon n}-1)a_n\bigg]$$
$$\leq\mathbb{P}\bigg[P_{\rm e}^{\rm ub}(\mathcal{C}_n)<2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\varepsilon/2)}\bigg]+2^{-n\big(\delta(R)-\varepsilon/2\big)}. \tag{81}$$

From (68), (71), and (81), we have

$$\mathbb{P}\bigg[\bigg|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\bigg|>\varepsilon\bigg]$$
$$\leq\mathbb{P}\bigg[P_{\rm e}^{\rm ub}(\mathcal{C}_n)>\frac{1}{2}2^{-n(E_{\rm trc}^{\rm ub}(R,Q)-\varepsilon/2)}\bigg]$$
$$+\mathbb{P}\bigg[P_{\rm e}^{\rm ub}(\mathcal{C}_n)<2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\varepsilon/2)}\bigg]+2^{-n\big(\delta(R)-\varepsilon/2\big)} \tag{82}$$
$$\leq\frac{1}{n^{1+\beta}}+2^{-n\big(\delta(R)-\varepsilon/2\big)} \tag{83}$$
$$\to 0, \tag{84}$$

for any $0<\varepsilon<2\delta(R)$, where (83) follows from Lemma 8 with β being a positive constant. Since $\mathbb{P}\big[\big|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\big|>\varepsilon\big]$ is a non-increasing function in ε, (84) must hold for all $\varepsilon>0$.

Furthermore, since $\mathbb{P}\big[\big|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\big|>\varepsilon\big]$ is a non-increasing function in ε, it holds that

$$\mathbb{P}\big[\big|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\big|>\varepsilon\big]\leq\frac{1}{n^{1+\beta}}+2^{-n\delta(R)/2} \tag{85}$$

for any $\varepsilon\in(0,2\delta(R))$. It follows from (85) that $\sum_{n=1}^{\infty}\mathbb{P}\big[\big|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\big|>\varepsilon\big]<\infty$. Hence, by Borel-Cantelli's lemma [24, Theorem 4.3], we have

$$E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\xrightarrow{\rm (a.s.)}0, \tag{86}$$

where $\xrightarrow{\rm (a.s.)}$ denotes almost sure convergence as $n\to\infty$, that is, a sequence of random variables $\{A_n\}_{n=1}^{\infty}$ converge almost surely to A if $\mathbb{P}[\lim_{n\to\infty}A_n=A]=1$. On the other hand, observe that

$$\big|E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\big|\leq-\frac{2\log P_{\rm e}(\mathcal{C}_n)}{n} \tag{87}$$
$$\leq 2E_{\rm sp}(R)+{\rm o}(1), \tag{88}$$

where (88) follows from the fact that the error exponent of any sufficiently long code is upper bounded by the sphere-packing bound $E_{\rm sp}(R)$ [4, Theorem 2]. Hence, from (86) and (88), by the bounded convergence theorem [24, Theorem 5.4], it holds that

$$\lim_{n\to\infty}\mathbb{E}\big[E_n(\mathcal{C}_n)-E_n^{\rm ub}(\mathcal{C}_n)\big]=0. \tag{89}$$

This means that

$$E_{\rm trc}(R,Q)=\lim_{n\to\infty}E_n^{\rm ub}(\mathcal{C}_n) \tag{90}$$
$$=E_{\rm trc}^{\rm ub}(R,Q). \tag{91}$$

2) Second term of (58)*:* Using Chebyshev's inequality, we have

$$\mathbb{P}\bigg[\bigg|E_n^{\rm ub}(\mathcal{C}_n)-\bigg(-\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]\bigg)\bigg|>\varepsilon\bigg]$$
$$\leq\frac{1}{\varepsilon^2}\,{\rm Var}\bigg(-\frac{\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n}\bigg). \tag{92}$$

Now, define $\xi(p,n,R)\triangleq 2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+R)}$.

From (92), we obtain

$$\mathbb{P}\bigg[\bigg|E_n^{\rm ub}(\mathcal{C}_n)-\bigg(-\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]\bigg)\bigg|>\varepsilon\bigg]$$
$$\leq\frac{1}{n^2\varepsilon^2}\mathbb{E}\bigg[\bigg(-\log(M_n-1)-\log\xi(p,n,R)$$
$$-\log\bigg(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\bigg)\bigg)^2\bigg]$$
$$-\frac{1}{\varepsilon^2}\bigg(\frac{\mathbb{E}\big[-\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]}{n}\bigg)^2. \tag{93}$$

By Lemma 7, we know that

$$\lim_{n\to\infty}\frac{\mathbb{E}\big[-\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]}{n}=E_{\rm trc}^{\rm ub}(R,Q), \tag{94}$$

hence, it holds that

$$\limsup_{n\to\infty}\mathbb{P}\left[\left|E_n^{\rm ub}(\mathcal{C}_n)-\left(-\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]\right)\right|>\varepsilon\right]$$
$$=\limsup_{n\to\infty}\frac{1}{\varepsilon^2}\mathbb{E}\Bigg[\bigg(E_{\rm trc}^{\rm ub}(R,Q) -\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\bigg)^2\Bigg]-\frac{\big(E_{\rm trc}^{\rm ub}(R,Q)\big)^2}{\varepsilon^2} \tag{95}$$
$$\le\frac{1}{\varepsilon^2}\Bigg((E_{\rm trc}^{\rm ub}(R,Q))^2-2E_{\rm trc}^{\rm ub}(R,Q) \times\liminf_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right] +\limsup_{n\to\infty}\mathbb{E}\left[\left(\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right)^2\right]\Bigg) -\frac{\big(E_{\rm trc}^{\rm ub}(R,Q)\big)^2}{\varepsilon^2}, \tag{96}$$

where (96) follows from the sub-additivity of $\limsup$. Now, we need to estimate

$$\liminf_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right]$$

and

$$\limsup_{n\to\infty}\mathbb{E}\left[\left(\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right)^2\right].$$

First, we show that

$$\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M-1)\xi(p,n,R)}\right)\xrightarrow{\text{(a.s.)}}0. \tag{97}$$

Indeed, take an arbitrary $\nu>0$ and observe that

$$\mathbb{P}\left[\left|\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right|>\nu\right]$$
$$=\mathbb{P}\left[\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{M_n-1}>2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+R-\nu)}\right] +\mathbb{P}\left[\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{M_n-1}<2^{-N(E_{\rm trc}^{\rm ub}(R,Q)+R+\nu)}\right] \tag{98}$$
$$\le\mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n)>\frac{1}{2}2^{-n(E_{\rm trc}^{\rm ub}(R,Q)-\nu)}\right] +\mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n)<2^{-n(E_{\rm trc}^{\rm ub}(R,Q)+\nu)}\right] \tag{99}$$
$$\le\frac{1}{n^{1+\beta}}, \tag{100}$$

for some constants $\beta>0$, where (100) follows from Lemma 8. From (100), we obtain

$$\sum_{n=1}^{\infty}\mathbb{P}\left[\left|\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right|>\nu\right] <\sum_{n=1}^{\infty}\frac{1}{n^{1+\beta}}<\infty \tag{101}$$

by using D'Alembert criterion. This means that (97) holds, or

$$\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\xrightarrow{\text{(a.s.)}}0 \tag{102}$$

by Borel-Cantelli lemma [24, Theorem 4.3]. Now, since $0\le\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)\le 1$ for all $i,j\in[M]:i\neq j$, it holds that

$$\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right) =\frac{1}{n}\log\left(\frac{1}{M_n(M_n-1)\xi(p,n,R)}\sum_{i\neq j}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)\right) \tag{103}$$
$$\le\frac{1}{n}\log\left(\frac{1}{\xi(p,n,R)}\right) \tag{104}$$
$$\le E_{\rm trc}^{\rm ub}(R,Q)+R, \tag{105}$$

where (105) follows from the definition of $\xi(p,n,R)$.On the other hand, from the sphere-packing bound [3], it holds almost surely that

$$\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right) \ge\frac{1}{n}\log\left(\frac{P_{\rm e}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right) \tag{106}$$
$$\dot{\ge}\frac{1}{n}\log\left(\frac{2^{-nE_{\rm sp}(R)}}{(M_n-1)\xi(p,n,R)}\right) \tag{107}$$
$$=E_{\rm trc}^{\rm ub}(R,Q)-E_{\rm sp}(R), \tag{108}$$

where (107) follows from the sphere-packing bound [4, Theorem 2], and (108) follows from the definition of $\xi(p,n,R)$ and $M_n=2^{nR}$.

From (105) and (108), $\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)$ is bounded (both below and above). Hence, by the bounded convergence theorem [24, Theorem 5.4] and the continuous mapping theorem [24, Theorem 4.3], it holds that

$$\mathbb{E}\left[\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right]\to 0, \tag{109}$$
$$\mathbb{E}\left[\left(\frac{1}{n}\log\left(\frac{P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{(M_n-1)\xi(p,n,R)}\right)\right)^2\right]\to 0. \tag{110}$$

From (96), (109), and (110), we finally have

$$\limsup_{n\to\infty}\mathbb{P}\left[\left|E_n^{\rm ub}(\mathcal{C}_n)-\left(-\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]\right)\right|>\varepsilon\right]=0 \tag{111}$$

for any arbitrary $\varepsilon>0$. This leads to

$$\lim_{n\to\infty}\mathbb{P}\left[\left|E_n^{\rm ub}(\mathcal{C}_n)-\left(-\frac{1}{n}\mathbb{E}\big[\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big]\right)\right|>\varepsilon\right]=0 \tag{112}$$

by the fact that the probability measure is bounded from below by zero.

3) Third term of (58)*:* By Lemma 7, it is known that

$$\mathbb{E}\left[\frac{-\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n}\right]\to E_{\rm trc}^{\rm ub}(R,Q). \tag{113}$$

On the other hand, from (91) in Step 1, we know that

$$E_{\rm trc}^{\rm ub}(R,Q)=E_{\rm trc}(R,Q). \tag{114}$$

[3]In case that the sphere packing bound diverges, we can use $E_{\rm ex}(R=0)$ as an upper bound, which is finite at $R=0$ unless the zero error capacity $C_0>0$.

It follows from (113) and (114) that

$$\mathbb{P}\left[\left|\mathbb{E}\left[\frac{-\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n}\right]-E_{\rm trc}(R,Q)\right|>\varepsilon\right]\to 0. \tag{115}$$

In conclusion, as anticipated, the three terms of (58) tend to zero as $n\to\infty$, showing (12) for rates below the critical rate. Together with the case $E_{\rm trc}(R,Q)=E_{\rm rce}(R,Q)$ in (42), we proved Theorem 1, which states the convergence in probability of the error exponent of the codes in the ensemble to the typical random-coding error exponent.

We end our proof with the extension to constant-composition codes. For the constant-composition code, and for all rates such that $R_{\rm crit}\le R\le C$, the proof of Theorem 1 holds by using the Levy's continuity theorem since it is not hard to see that $E_{\rm rce}(R,Q)=E_{\rm trc}(R,Q)$ for this case. At all the rate $0\le R\le E_{\rm trc}(R,Q)$, Lemma 4 - Lemma 6 still hold since $\mathbb{1}\{(\boldsymbol{X}_i,\boldsymbol{X}_j)\in\mathcal{T}_n(Q_{XX'})\}$ and $\mathbb{1}\{(\boldsymbol{X}_k,\boldsymbol{X}_l)\in\mathcal{T}_n(\tilde{Q}_{XX'})\}$ are still pairwise-independent for the constant-composition code for all $\{i,j,k,l\in[M]:i\neq j,k\neq l\}$. In Lemma 7, the typical error exponent of the union bound should be replaced by $E_{\rm trc}^{\rm ub}$ for the constant-composition code in [9, Theorem 1]. To show that Theorem 1 still holds for the constant-composition code, we need to prove that the mapping from the error probability and the union bound in Lemma 8 and Lemma 9 still work. It is not hard to see that the proof of Lemma 8 still holds for the constant-composition code since its correctness depends on Lemma 4, Lemma 6 and the fact that $\tilde{V}_{ij}$'s are pairwise-independent where $\tilde{V}_{ij}$ is defined in (450). Lemma 9 still holds for the constant-composition code, as stated as follows.

Lemma 10: For any constant-composition code with type Q and for all rates such that $0<R<R_{\rm crit}$, we have

$$0\le\frac{\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]}{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]}-1\le 2^{-n\left(\delta(R)+E_{\rm trc}^{\rm ub}(R,Q)-E_{\rm rce}(R,Q)\right)} \tag{116}$$

for some $\delta(R)>0$.

Proof: To prove Lemma 10, we use the same proof as Lemma 9 in Appendix A.8. In fact, equation (520) still holds for the constant-composition code. In addition, the pairwise error probability only depends on the joint-type of the two codewords as in the i.i.d .case. ■

B. Proof of Theorem 2

The proof of Theorem 2 is structured as follows. From (117) to (150) we first prove (13), and then from (151) to (156) we prove (14), both for the i.i.d. ensemble.

To prove (13), Under the condition that $E_{\rm rce}(R,Q)=E_{\rm trc}(R,Q)$, we observe that

$$\begin{aligned}&\mathbb{P}\left[-\frac{1}{n}\log P_{\rm e}(\mathcal{C}_n)<E_{\rm trc}(R,Q)-\varepsilon\right]\\ &=\mathbb{P}\left[P_{\rm e}(\mathcal{C}_n)>2^{-n(E_{\rm trc}(R,Q)-\varepsilon)}\right] &(117)\\ &\doteq 2^{n(E_{\rm trc}(R,Q)-\varepsilon)}2^{-nE_{\rm rce}(R,Q)} &(118)\\ &=2^{-n\varepsilon}, &(119)\end{aligned}$$

where (118) follows from Markov's inequality and $\mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)\big]\doteq 2^{-nE_{\rm rce}(R,Q)}$, (119) follows from $E_{\rm rce}(R,Q)=E_{\rm trc}(R,Q)$. Now, for any $s>0$, observe that

$$\begin{aligned}&\mathbb{P}\left[-\frac{1}{n}\log P_{\rm e}(\mathcal{C}_n)<E_{\rm trc}(R,Q)-\varepsilon\right]\\ &=\mathbb{P}\left[2^{\frac{s}{n}\log P_{\rm e}(\mathcal{C}_n)}>2^{-s(E_{\rm trc}(R,Q)-\varepsilon)}\right] &(120)\\ &\le 2^{s(E_{\rm trc}(R,Q)-\varepsilon)}\mathbb{E}\left[2^{\frac{s}{n}\log P_{\rm e}(\mathcal{C}_n)}\right] &(121)\\ &\le 2^{s(E_{\rm trc}(R,Q)-\varepsilon)}\mathbb{E}\big[\big(P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big)^{s/n}\big]. &(122)\end{aligned}$$

On the other hand, for any $0\le s\le n$ and $\lambda>0$, we have

$$\begin{aligned}&\mathbb{E}\big[\big(P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big)^{s/n}\big]\\ &=\mathbb{E}\left[\left(\frac{1}{M}\sum_{i\neq j}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)\right)^{s/n}\right] &(123)\\ &\le\frac{1}{M^{s/n}}\sum_{i\neq j}\mathbb{E}\left[\Big(\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)\Big)^{s/n}\right] &(124)\\ &=\frac{M(M-1)}{M^{s/n}}\mathbb{E}\left[\Big(\mathbb{P}(\boldsymbol{X}_1\to\boldsymbol{X}_2)\Big)^{s/n}\right], &(125)\end{aligned}$$

where (124) follows from $(x_1+x_2+\cdots+x_n)^\alpha\le x_1^\alpha+x_2^\alpha+\cdots+x_n^\alpha$ for any $x_1,x_2,\cdots,x_n\ge 0$ while $\alpha\in[0,1]$.

On the other hand, by Lemma 4, the probability $\mathbb{P}(\boldsymbol{X}_1\to\boldsymbol{X}_2)$ with joint type $Q_{XX'}$ satisfies

$$\mathbb{P}(\boldsymbol{X}_1\to\boldsymbol{X}_2)=2^{-n\sum_{k=1}^n\sum_{x,x'}d_{\rm B}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}}. \tag{126}$$

Hence, for any $0\le s\le n$, we have

$$\begin{aligned}&\mathbb{E}\left[\big(\mathbb{P}(\boldsymbol{X}_1\to\boldsymbol{X}_2)\big)^{\frac{s}{n}}\right]\\ &=\mathbb{E}\left[2^{\frac{s}{n}\log\mathbb{P}(\boldsymbol{X}_1\to\boldsymbol{X}_2)}\right] &(127)\\ &=\mathbb{E}\left[2^{-\frac{s}{n}\sum_{k=1}^n\sum_{x,x'}d_{\rm B}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}}\right]. &(128)\end{aligned}$$

Now, since $\{\sum_{x,x'}d_{\rm B}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}\}_{k=1}^n$ are i.i.d., by the SLLN, we have

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^n\sum_{x,x'}d_{\rm B}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}\\ &\xrightarrow{\text{(a.s.)}}\sum_{x,x'}Q(x)Q(x')d_{\rm B}(x,x'). &(129)\end{aligned}$$

On the other hand, we have

$$\begin{aligned}0&\le\frac{1}{n}\sum_{k=1}^n\sum_{x,x'}d_{\rm B}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}\\ &\le\max_{x,x'}d_{\rm B}(x,x')<\infty. &(130)\end{aligned}$$

Hence, by the bounded convergence theorem [24, Theorem 5.4], we have

$$\begin{aligned}&\mathbb{E}\left[2^{-\frac{s}{n}\sum_{k=1}^n\sum_{x,x'}d_{\rm B}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}}\right]\\ &\to 2^{-s\sum_{x,x'}Q(x)Q(x')d_{\rm B}(x,x')}. &(131)\end{aligned}$$

Similarly, for any fixed constant $\lambda \geq 0$, we have

$$\mathbb{E}\left[2^{-\frac{s}{n(1+\lambda)}\sum_{k=1}^{n}\sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}}\right] \to 2^{-\frac{s}{1+\lambda}\sum_{x,x'} Q(x)Q(x')d_{\mathrm{B}}(x,x')}. \tag{132}$$

Now, let

$$J_k := \sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}. \tag{133}$$

Then, from (131) and (132), for any fixed constant $\lambda \geq 0$, it holds that

$$\mathbb{E}\left[2^{-\frac{s}{n}\sum_{k=1}^{n}\sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}}\right]$$
$$= (1+o(1))\left(\mathbb{E}\left[2^{-\frac{s}{n(1+\lambda)}\sum_{k=1}^{n} J_k}\right]\right)^{1+\lambda} \tag{134}$$
$$= (1+o(1))\left(\mathbb{E}\left[2^{-\frac{s}{n(1+\lambda)} J_1}\right]\right)^{n(1+\lambda)}. \tag{135}$$

From (125) and (135), we obtain

$$\mathbb{E}\left[\left(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)\right)^{s/n}\right] \leq (1+o(1))M^{2-\frac{s}{n}}\left(\mathbb{E}\left[2^{-\frac{s}{n(1+\lambda)} J_1}\right]\right)^{n(1+\lambda)}. \tag{136}$$

Now, observe that

$$\mathbb{E}\left[2^{-\frac{s}{n(1+\lambda)} J_1}\right] = \sum_{x,x'} \mathbb{P}((X_{11},X_{21})=(x,x')) \times \mathbb{E}\left[2^{-\frac{s}{n(1+\lambda)} J_1}\middle|(X_{11},X_{21})=(x,x')\right] \tag{137}$$
$$= \sum_{x,x'} Q(x)Q(x')2^{-\frac{s}{n(1+\lambda)}d_{\mathrm{B}}(x,x')}. \tag{138}$$

From (136) and (138), we obtain

$$\mathbb{E}\left[\left(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)\right)^{\frac{s}{n}}\right] \leq (1+o(1))M^{2-\frac{s}{n}}\left(\sum_{x,x'} Q(x)Q(x')2^{-\frac{s}{n(1+\lambda)}d_{\mathrm{B}}(x,x')}\right)^{n(1+\lambda)}. \tag{139}$$

From (122) and (139), for any s such that $0 \leq s \leq n$ and any fixed constant $\lambda > 0$, we have

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) < E_{\mathrm{trc}}(R,Q)-\varepsilon\right] \leq (1+o(1))2^{s(E_{\mathrm{trc}}(R,Q)-\varepsilon)}M^{2-\frac{s}{n}} \times \left(\sum_{x,x'} Q(x)Q(x')2^{-\frac{s}{n(1+\lambda)}d_{\mathrm{B}}(x,x')}\right)^{n(1+\lambda)}. \tag{140}$$

From (140), by choosing $s=n$ and using $M=2^{nR}$, we have

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) < E_{\mathrm{trc}}(R,Q)-\varepsilon\right] \leq (1+o(1))2^{n\left[(1+\lambda)\log\left(\sum_{x,x'} Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda}}\right)\right]} \times 2^{n(E_{\mathrm{trc}}(R,Q)+R-\varepsilon)}. \tag{141}$$

Now, for $E_{\mathrm{trc}}(R,Q) \neq E_{\mathrm{rce}}(R,Q)$, from (91) and Lemma 7, observe that

$$E_{\mathrm{trc}}(R,Q) = \min_{P_{XX'}:D(P_{XX'}\|Q_XQ_{X'})\leq 2R} D(P_{XX'}\|Q_XQ_{X'}) + \sum_{x,x'} d_{\mathrm{B}}(x,x')P_{XX'}(x,x') - R. \tag{142}$$

Given the distribution Q and $Q_X = Q_{X'} = Q$, the optimization problem in (142) is convex in $\{P_{XX'}(x,x')\}_{x,x'}$ since the KL divergence is convex. By using standard Karush-Kuhn-Tucker conditions, it can be seen that (142) has as two optimal solutions $P_{XX'} \in \{P^0_{XX'}, P^*_{XX'}\}$ given by

$$P^0_{XX'}(x,x') = \frac{Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}}{\sum_{x,x'} Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}}, \tag{143}$$

$$P^*_{XX'}(x,x') = \frac{Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda_*}}}{\sum_{x,x'} Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda_*}}}, \tag{144}$$

where λ^* is the unique positive solution of $2R = D(P^*_{XX'}(x,x')\|Q_XQ_{X'})$. For $P_{XX'} = P^0_{XX'}$, we obtain, after some algebra, that the following terms in the exponent of the r.h.s. of (141) vanish. More specifically, that

$$E_{\mathrm{trc}}(R,Q) + R + \log\left(\sum_{x,x'} Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}\right)$$
$$= D(P^0_{XX'}\|Q_XQ_{X'}) + \sum_{x,x'} d_{\mathrm{B}}(x,x')P^0_{XX'}(x,x') + \log\left(\sum_{x,x'} Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}\right) \tag{145}$$
$$= \sum_{x,x'} \frac{Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}}{\sum_{x,x'} Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}} \times \log\frac{2^{-d_{\mathrm{B}}(x,x')}}{\sum_{x,x'} Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}} + \sum_{x,x'} d_{\mathrm{B}}(x,x')P^0_{XX'}(x,x') + \log\left(\sum_{x,x'} Q(x)Q(x')2^{-d_{\mathrm{B}}(x,x')}\right) \tag{146}$$
$$= 0. \tag{147}$$

For the case $P_{XX'} = P^*_{XX'}$, by performing similarly manipulations we obtain that

$$E_{\mathrm{trc}}(R,Q) + R + (1+\lambda^*)\log\left(\sum_{x,x'} Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda^*}}\right)$$
$$\leq (1+\lambda^*)\sum_{x,x'} \frac{Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda^*}}}{\sum_{x,x'} Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda^*}}} \times \log\frac{2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda^*}}}{\sum_{x,x'} Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda^*}}} + \sum_{x,x'} d_{\mathrm{B}}(x,x')P^*_{XX'}(x,x') + (1+\lambda^*)\log\left(\sum_{x,x'} Q(x)Q(x')2^{-\frac{d_{\mathrm{B}}(x,x')}{1+\lambda^*}}\right) \tag{148}$$

$$= 0, \tag{149}$$

The results in (147), and (149), after choosing $\lambda = 0$ for the first case and $\lambda = \lambda^*$ for the second case, used in (141), imply that

$$\mathbb{P}\bigg[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) < E_{\mathrm{trc}}(R,Q) - \varepsilon\bigg] \doteq 2^{-n\varepsilon}. \tag{150}$$

Finally, from (119) and (150), we obtain (13), concluding our proof for the i.i.d. random codebook ensemble.

We finally prove (14). For any i.i.d. code the following holds:

$$P_{\mathrm{e}}(\mathcal{C}_n) \ge \max_{i\neq j} \mathbb{P}(\boldsymbol{X}_i \to \boldsymbol{X}_j). \tag{151}$$

Defining $V_n = -\frac{1}{n}\log \max_{i\neq j}\mathbb{P}(\boldsymbol{X}_i \to \boldsymbol{X}_j)$, and setting

$$K_1 = \sum_{x,x'} d_{\mathrm{B}}^2(x,x')Q(x)Q(x'), \tag{152}$$

$$K_2 = \sum_{x,x'} d_{\mathrm{B}}(x,x')Q(x)Q(x'). \tag{153}$$

It follows from (151) that

$$\mathbb{P}\bigg[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) > E_{\mathrm{trc}}(R,Q) + \varepsilon\bigg]$$

$$\le \mathbb{P}\Bigg[\frac{V_n - \mathbb{E}[V_n]}{\sqrt{\mathrm{Var}(V_n)}} > \frac{n\big(E_{\mathrm{trc}}(R,Q) + \varepsilon - \frac{\mathbb{E}[V_n]}{n}\big)}{\sqrt{\mathrm{Var}(V_n)}}\Bigg] \tag{154}$$

$$= Q\Bigg(\frac{n\big(E_{\mathrm{trc}}(R,Q) + \varepsilon - \frac{\mathbb{E}[V_n]}{n}\big)}{\sqrt{\mathrm{Var}(V_n)}}\Bigg) + O\bigg(\frac{1}{\sqrt{n}}\bigg) \tag{155}$$

$$= Q\Bigg(\frac{\sqrt{n}\big(E_{\mathrm{trc}}(R,Q) + \varepsilon - E_{\mathrm{trc}}(Q,0)\big)}{\sqrt{K_1 - K_2^2}}\Bigg) + O\bigg(\frac{1}{\sqrt{n}}\bigg) \tag{156}$$

as $n \to \infty$ since $E_{\mathrm{trc}}(R,Q) \ge E_{\mathrm{trc}}(0,Q)$, where (155) follows as a sub-result from the proof of Theorem 6, and (156) follows from (353) and the Berry–Esseen theorem [16, Theorem 3.4.17].

C. Proof of Theorem 3

We start the proof of Theorem 3 with some auxiliary results until (159), and then discuss the two terms in (160): the first term from (162) to (178), and the second term from (179) to (187).

Lemma 11: Let $\mathcal{I}\{i,j\} = \mathcal{I}\{(\mathbf{x}_i,\mathbf{x}_j) \in \mathcal{T}(P_{XX'})\}$, where $\mathcal{I}\{.\}$ is the indicator function. Then, for $0 \le \eta \le D(P_{XX'}\|Q_XQ'_X)$, it holds that $2^{-n2D(P_{XX'}\|Q_XQ'_X)} \le \mathbb{E}[\mathcal{I}\{i,j\}\mathcal{I}\{i,k\}] \le 2^{-n[D(P_{XX'}\|Q_XQ'_X)+\eta]}$. ∎

Proof: Appendix A.9. ∎

Lemma 12: For any $\epsilon > 0$ and for any joint type $P_{XX'}$ such that $D(P_{XX'}\|Q_XQ'_X) \le R - \epsilon$, $\forall \epsilon > 0$, the following holds:

$$\mathbb{P}\left[\mathcal{N}(P_{XX'}) \le 2^{-n\epsilon}\mathbb{E}[\mathcal{N}(P_{XX'})]\right] \dot{\le} 2^{-2^{n\epsilon}} \tag{157}$$

Proof: Appendix A.10. ∎

Using Lemma 11 and Lemma 12 we prove the following theorem, which states that the probability of finding a code for which the exponent of $P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)$ is larger than the expurgated exponent $E_{\mathrm{ex}}(R)$ is double-exponentially decaying in n. Now we can prove the main part of theorem 3. We have that

$$P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) \doteq \max_{P_{XX'}} \mathcal{N}(P_{XX'})e^{-n[\sum_{x,x'} d_{\mathrm{B}}(x,x')P_{XX'}(x,x')+R]}. \tag{158}$$

Let us refer to the maximizing joint type of (158) as $P^*_{XX'}$. We define the following complementary events:

$$A = \{P^*_{XX'} \in \mathcal{P}\}\ ,\quad \overline{A} = \{P^*_{XX'} \in \overline{\mathcal{P}}\} \tag{159}$$

where $\mathcal{P} = \{P_{XX'}|D(P_{XX'}\|Q_XQ'_X) \le 2R\}$, $Q_XQ'_X$ being the theoretical joint type, while $\overline{\mathcal{P}}$ is the complement to set $\mathcal{P}$. Consider a positive real number $E_2 > E_{\mathrm{trc}}(R,Q)$. We have:

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) \ge E_2\right]$$

$$= \mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) \ge E_2, A\right]$$

$$+ \mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) \ge E_2, \overline{A}\right]. \tag{160}$$

Now we proceed to bound from above both terms at the right hand side of (160). Define

$$F(P_{XX'}) = \sum_{x,x'} d_{\mathrm{B}}(x,x')P_{XX'}(x,x'). \tag{161}$$

1) First Term:

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) \ge E_2, A\right]$$

$$= \mathbb{P}\left[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) \le 2^{-nE_2}, A\right] \tag{162}$$

$$= \mathbb{P}\left[\frac{1}{M_n}\sum_{P_{XX'}} \mathcal{N}(P_{XX'})2^{-nF(P_{XX'})} \le 2^{-nE_2}, A\right] \tag{163}$$

$$\doteq \mathbb{P}\left[\max_{P_{XX'}} \mathcal{N}(P_{XX'})2^{-nF(P_{XX'})} \le 2^{-n(E_2-R)}, A\right] \tag{164}$$

$$\le \mathbb{P}\left[\max_{P_{XX'}\in\mathcal{P}} \mathcal{N}(P_{XX'})2^{-nF(P_{XX'})} \le 2^{-n(E_2-R)}\right] \tag{165}$$

$$= \mathbb{P}\left[\bigcap_{P_{XX'}\in\mathcal{P}} \left[\mathcal{N}(P_{XX'}) \le 2^{-n(E_2-R-F(P_{XX'})}\right]\right\} \tag{166}$$

where (165) follows from the definition of A and from removing the event A. Let us now define $\mathcal{P}'$:

$$\mathcal{P}' = \{P_{XX'}|D(P_{XX'}\|Q_XQ'_X) \le R\}, \tag{167}$$

and note that $\mathcal{P}' \subset \mathcal{P}$. Let us consider the term $2^{-n(E_2-R-F(P_{XX'}))}$. We now look for a $P_{XX'} \in \mathcal{P}'$ such that this is smaller than the mean of the enumerator function, i.e, a $P_{XX'} \in \mathcal{P}'$ such that the following holds:

$$2^{-n(E_2-R-F(P_{XX'})} \le 2^{n[2R-D(P_{XX'}\|Q_XQ'_X)-\epsilon]} \tag{168}$$

$$E_2 \ge -R + D(P_{XX'}\|Q_XQ'_X) + F(P_{XX'}) + \epsilon. \tag{169}$$

Let us indicate the $P_{XX'}$ that minimizes (169) with $P'_{XX'}$. Minimizing the term at the right hand side of (169) we can set the value of E_2 to:

$$E_2 = \min_{P_{XX'}\in\mathcal{P}'} -R + D(P_{XX'}\|Q_XQ'_X) + F(P_{XX'}) + \epsilon. \tag{170}$$

The right hand side of (170) is strictly larger than $E_{\rm trc}(R,Q)$. To see this note the following:

$$\min_{P_{XX'}\in\mathcal{P}'} -R + D(P_{XX'}\|Q_XQ'_X) + F(P_{XX'}) + \epsilon \tag{171}$$
$$> \min_{P_{XX'}\in\mathcal{P}} R - 2R + D(P_{XX'}\|Q_XQ'_X) + F(P_{XX'}) + \epsilon \tag{172}$$
$$= \min_{P_{XX'}\in\mathcal{Z}_{GGV}} R - 2R + D(P_{XX'}\|Q_XQ'_X) + F(P_{XX'}) + \epsilon \tag{173}$$
$$= \min_{P_{XX'}\in\mathcal{Z}_{GGV}} R + F(P_{XX'}) + \epsilon \tag{174}$$
$$= E_{\rm trc}(R,Q) + \epsilon \tag{175}$$

where (172) follows from the fact that $\mathcal{P}'\subset\mathcal{P}$, (173) follows from the concavity of the objective function (minimum is on the border) while (175) follows from the definition of $\mathcal{Z}_{GGV}$. With this definition of E_2 we ensure that for at least one joint type the conditions for applying Lemma 12 (i.e., (168)) hold. Using the definition in (166) together with the statement of Lemma 12 we have:

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\rm e}^{\rm ub}(\mathcal{C}_n) \geq E_2, A\right]$$
$$\dot{\leq} \mathbb{P}\left[\bigcap_{P_{XX'}\in\mathcal{P}}\left[\mathcal{N}(P_{XX'}) \leq 2^{-n(E_2-R+F(P_{XX'})}\right]\right\} \tag{176}$$
$$\dot{\leq} 2^{-2^{n[R-D(P'_{XX'}\|Q_XQ'_X)]}} \tag{177}$$
$$\leq 2^{-2^{n\epsilon'}} \tag{178}$$

with $\epsilon'>0$.

2) Second Term:

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\rm e}^{\rm ub}(\mathcal{C}_n) \geq E_2, \overline{A}\right]$$
$$= \mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n) \leq 2^{-nE_2}, \overline{A}\right] \tag{179}$$
$$= \mathbb{P}\left[\frac{1}{M_n}\sum_{P_{XX'}}\mathcal{N}(P_{XX'})2^{nF(P_{XX'})} \leq 2^{-nE_2}, \overline{A}\right] \tag{180}$$
$$\doteq \mathbb{P}\left[\max_{P_{XX'}}\mathcal{N}(P_{XX'})2^{nF(P_{XX'})} \leq 2^{-n(E_2-R)}, \overline{A}\right]. \tag{181}$$

Consider (181). The event $\overline{A}$ implies that the joint type maximizing the expression at the left hand side lays outside $\mathcal{P}$. This implies that any $P_{XX'}$ which lies inside $\mathcal{P}$ leads to a value which is no greater than the maximum. Since this is an implication of the events within brackets, its probability is larger than or equal to the one of (181).Thus we have:

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\rm e}^{\rm ub}(\mathcal{C}_n) \geq E_2, \overline{A}\right]$$
$$\doteq \mathbb{P}\left[\max_{P_{XX'}}\mathcal{N}(P_{XX'})2^{nF(P_{XX'})} \leq 2^{-n(E_2-R)}, \overline{A}\right] \tag{182}$$
$$\leq \mathbb{P}\left[\max_{P_{XX'}\in\mathcal{P}}\mathcal{N}(P_{XX'})2^{nF(P_{XX'})} \leq 2^{-n(E_2-R)}\right] \tag{183}$$
$$\leq 2^{-2^{n\epsilon'}} \tag{184}$$

where (184) is because (182) has the same form as (165) and thus the same inequalities as for the first term hold.

Finally, we note that from (170) we can further state the following:

$$E_2 = \min_{P_{XX'}\in\mathcal{P}'} -R + D(P_{XX'}\|Q_XQ'_X) + F(P_{XX'}) + \epsilon \tag{185}$$
$$= \min_{P_{XX'}\in\mathcal{P}'} -\sum_{x,x'} d_{\rm B}(x,x')P_{XX'}(x,x') + \epsilon \tag{186}$$
$$= E_{\rm ex}(R) + \epsilon \tag{187}$$

where (186) follows from the concavity of the objective function, which implies that the minimum is on the border of the region $\mathcal{P}'$, and from the definition of $\mathcal{P}'$ while (187) is found by calculating the derivative of [22, Eq. (5.7.11)] with respect to the optimization variable ρ and, after some change of variable, equating to zero.

D. *Proof of Theorem 4*

Now let us consider the following inequality

$$P_{\rm e}^{\rm ub}(\mathcal{C}_n) \leq M_n P_{\rm e}(\mathcal{C}_n) \tag{188}$$

which follows from upper-bounding the probability $\mathbb{P}[\boldsymbol{x}_i\to\boldsymbol{x}_j]$ in (15) by $\mathbb{P}\left[\bigcup_{j\neq i}\{\boldsymbol{x}_i\to\boldsymbol{x}_j\}\right]$ in (4). From Theorem 3 and using (188) we have

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\rm e}^{\rm ub}(\mathcal{C}_n) \geq E_{\rm ex}(R) + R + \epsilon\right] \leq 2^{-2^{n\epsilon}} \tag{189}$$

and finally (18).

E. *Proof of Theorem 5*

This proof is split into two parts, the first part from (192) to (219) is devoted to the i.i.d. ensemble, while the second part from (220) to (264) deals with the constant-composition ensemble.

1) i.i.d. ensemble: Observe that

$$\max_{i\neq j}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)$$
$$\leq P_{\rm e}(\mathcal{C}_n) \tag{190}$$
$$\leq \sum_{i=1}^{M_n}\sum_{j\neq i}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j) \tag{191}$$
$$\leq M_n(M_n-1)\max_{i\neq j}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j). \tag{192}$$

On the other hand, by Lemma 4, the pairwise codeword error probability $\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)$ given $P_{XX'}$ satisfies

$$\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j) = 2^{-n\sum_{x,x'}d_{\rm B}(x,x')\hat{P}_{\boldsymbol{X}_i\boldsymbol{X}_j}(x,x')}, \tag{193}$$

where $\hat{P}_{\boldsymbol{X}_i\boldsymbol{X}_j}$ is the n-joint type of $(\boldsymbol{X}_1,\boldsymbol{X}_2)$. Observe that

$$\hat{P}_{\boldsymbol{X}_i\boldsymbol{X}_j}(x,x') = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}. \tag{194}$$

It follows from (193) and (194) that

$$\mathbb{P}(\boldsymbol{X}_i \to \boldsymbol{X}_j) = 2^{-\sum_{k=1}^n \sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}} \tag{195}$$

for all $i,j \in [M_n], i \neq j$. Since M_n sub-exponential in n, from (192) and (195), we obtain

$$-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \sim \frac{V_n}{n}, \tag{196}$$

where $X \sim Y$ means that X and Y have the same asymptotic distributions, and

$$V_n = \min_{i\neq j} Z_{ij} \tag{197}$$

with

$$Z_{ij} = \sum_{k=1}^n \sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}, \tag{198}$$

for all $i,j \in [M_n]$ and $i \neq j$. Now, observe that

$$\mathbb{E}[Z_{ij}] = \sum_{k=1}^n \sum_{x,x'} d_{\mathrm{B}}(x,x')Q(x)Q(x'). \tag{199}$$

In addition, we have, after some algebra, that

$$\begin{aligned}\mathrm{Var}(Z_{ij}) = n\bigg(&\sum_{x,x'} d_{\mathrm{B}}^2(x,x')Q(x)Q(x') \\ &- \bigg(\sum_{x,x'} d_{\mathrm{B}}(x,x')Q(x)Q(x')\bigg)^2\bigg). \end{aligned}\tag{200}$$

for all $i \neq j$. Now, define

$$\begin{aligned} T_{ij} &:= \frac{Z_{ij}-\mathbb{E}[Z_{ij}]}{\sqrt{\mathrm{Var}(Z_{ij})}} &(201)\\ &= \frac{Z_{ij}-\mathbb{E}[Z_{12}]}{\sqrt{\mathrm{Var}(Z_{12})}}, &(202)\end{aligned}$$

where (202) follows from the fact that Z_{ij}'s are identically distributed. Then, by CLT, it holds that

$$T_{ij} \xrightarrow{(\mathrm{d})} \mathcal{N}(0,1), \qquad \forall i \neq j. \tag{203}$$

On the other hand, let

$$\begin{aligned}&\Upsilon_{ij}(k) \\ &= \sum_{x,x'} d_{\mathrm{B}}(x,x')\big(\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\} - Q(x)Q(x')\big).\end{aligned} \tag{204}$$

Then, for any fixed tuple $(\{\alpha_{ij}\} : i,j \in [M], i \neq j)$, we have

$$\begin{aligned}\sum_{i\neq j}\alpha_{ij}T_{ij} &= \frac{\sum_{i\neq j}\alpha_{ij}\sum_{k=1}^N \Upsilon_{ij}(k)}{\sqrt{\mathrm{Var}(Z_{12})}} &(205)\\ &= \sum_{k=1}^N \frac{\sum_{i\neq j}\alpha_{ij}\Upsilon_{ij}(k)}{\sqrt{\mathrm{Var}(Z_{12})}}. &(206)\end{aligned}$$

Now, by the i.i.d. random codebook generation, it holds that $\{\bar{V}_k\}_{k=1}^n$ are i.i.d. random variables, where

$$\bar{V}_k = \frac{\sum_{i\neq j}\alpha_{ij}\Upsilon_{ij}(k)}{\sqrt{\mathrm{Var}(Z_{12})}}. \tag{207}$$

In addition, since $(X_{i1},X_{j1})_{i\neq j}$'s are pairwise independent, we have

$$\mathrm{Var}(\bar{V}_1) = \frac{\sum_{i\neq j}\alpha_{ij}^2}{n}. \tag{208}$$

Hence, it holds from (206) and (208) that

$$\begin{aligned}\sum_{i\neq j}\alpha_{ij}T_{ij} &= \sqrt{\sum_{i\neq j}\alpha_{ij}^2}\bigg(\frac{\sum_{k=1}^n \bar{V}_k}{\sqrt{n\,\mathrm{Var}(\bar{V}_1)}}\bigg) \\ &\xrightarrow{(\mathrm{d})} \mathcal{N}\bigg(0, \sum_{i\neq j}\alpha_{ij}^2\bigg),\end{aligned} \tag{209}$$

where (209) follows from the CLT. Hence, the distribution of the vector $\{T_{ij} : i,j \in [M], i \neq j\}$ goes to the distribution of a jointly Gaussian random vector by the Levy's continuity theorem [24, Theorem 26.3].

Now, it is known that the distribution of any Gaussian random vector (both p.d.f and c.d.f.) is defined by its mean and covariance matrix. Since the covariance matrix of the vector $\{T_{ij} : i,j \in [M], i \neq j\}$ is the identity matrix by the pairwise independence of T_{ij}, which originates from the pairwise independence of $\mathbb{P}(\boldsymbol{X}_i \to \boldsymbol{X}_j)'s$, hence, the limit distribution is the standard normal Gaussian vector with dimension $M(M-1)$. This distribution is equal to the joint distribution of $M(M-1)$ independent standard normal variables $\{U_{ij}\}_{i\neq j}$. Hence, by the continuous mapping theorem [24, Theorem 25.6], it follows that

$$\min_{i\neq j} T_{ij} \xrightarrow{(\mathrm{d})} \min_{i\neq j} U_{ij}. \tag{210}$$

Now, observe that

$$\begin{aligned}\mathbb{E}\big[\big|\min_{i\neq j} T_{ij}\big|^4\big] &\leq \mathbb{E}\bigg[\bigg|\sum_{i\neq j}|T_{ij}|\bigg|^4\bigg] &(211)\\ &\leq \bigg(\sum_{i\neq j}1^{4/3}\bigg)^3\bigg(\sum_{i\neq j}\mathbb{E}\big[|T_{ij}|^4\big]\bigg) &(212)\\ &= M^4(M-1)^4\frac{\mathbb{E}\big[|Z_{12}-\mathbb{E}[Z_{12}|^4\big]}{\mathrm{Var}(Z_{12})^2} &(213)\\ &\leq 8M^4(M-1)^4\frac{\mathbb{E}\big[|Z_{12}|^4\big]}{\mathrm{Var}(Z_{12})^2}, &(214)\end{aligned}$$

where (212) and (214) follow from Hölder's inequality for the counting measure [28, Sec. 7.2].

Now, by (198), we have

$$Z_{12} = \sum_{k=1}^n \sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{1k},X_{2k})=(x,x')\}, \tag{215}$$

which is the sum of n independent random variables. Hence, we have

$$\mathbb{E}[|Z_{12}|^4] = O(n^2), \qquad \mathrm{Var}(Z_{12}) = \Theta(n). \tag{216}$$

Hence, from (214), we have

$$\mathbb{E}\big[\big|\min_{i\neq j} T_{ij}\big|^4\big] \leq \mathbb{E}\bigg[\bigg|\sum_{i\neq j}|T_{ij}|\bigg|^4\bigg] = O(1). \tag{217}$$

Hence, by Lemma 1 with $\varepsilon = 2$ and (210), we have

$$\frac{\min_{i\neq j} T_{ij} - \mathbb{E}[\min_{i\neq j} T_{ij}]}{\sqrt{\mathrm{Var}\left(\min_{i\neq j} T_{ij}\right)}} \xrightarrow{\text{(d)}} \frac{\min_{i\neq j} U_{ij} - \mathbb{E}[\min_{i\neq j} U_{ij}]}{\sqrt{\mathrm{Var}\left(\min_{i\neq j} U_{ij}\right)}}. \tag{218}$$

From (202), (218) and (196) we obtain (20), i.e.,

$$\frac{E_n(\mathcal{C}_n) - \mathbb{E}[E_n(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(E_n(\mathcal{C}_n))}} \xrightarrow{\text{(d)}} \frac{\min_{i\neq j} U_{ij} - \mathbb{E}[\min_{i\neq j} U_{ij}]}{\sqrt{\mathrm{Var}(\min_{i\neq j} U_{ij})}}. \tag{219}$$

2) Constant-composition ensemble: In this part, we use Stein's method to derive some criteria that provide sufficient conditions for the convergence in distribution to the normal random variable of the error probabilities and error exponents for general random coding ensemble over general channels, including the zero rate where $M_n \to \infty$ as we mentioned. This includes other random codebooks than i.i.d. random codebook ensembles.

We start by showing that Theorem 5 also holds for the constant-composition codes. In order to do this, we need some extra lemmas. First, we show the following fact which is based our modification of the Stein's criteria in [25, Theorem 3.2] to accommodate for the dependence among the random variables in the following lemma.

Lemma 13: Let $X_1, X_2, \cdots, X_n$ be zero-mean random variables on some alphabet $\mathcal{X} \subset \mathbb{R}$ such that $\sum_{i=1}^n \mathbb{E}[X_i^2] = n$. In addition, assume there exist positive sequences $\{\xi_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$ and a set $\mathcal{V} \subset \mathbb{R}^n$ with cardinality $|\mathcal{V}|$ such that

$$\begin{aligned}&\big(1-\xi_n\big)\prod_{i=1}^n \mathbb{P}[X_i = x_i] \\ &\le \mathbb{P}[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n] \\ &\le \big(1+\xi_n\big)\prod_{i=1}^n \mathbb{P}[X_i = x_i]\end{aligned} \tag{220}$$

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

$$\begin{aligned}\max\bigg\{&1, \frac{1}{\sqrt{n}}\sum_{i=1}^n |x_i|, \frac{1}{n}\bigg(\sum_{i=1}^n x_i\bigg)^2, \\ &\frac{1}{n}\sum_{i=1}^n x_i^2, \frac{1}{n^{3/2}}\sum_{i=1}^n |x_i|^3\bigg\} \le g_n, \forall (x_1, \cdots, x_n) \in \mathcal{V}^c.\end{aligned} \tag{221}$$

Assume also that $g_n \xi_n \to 0$ and

$$g_n \max\{\mathbb{P}(V^c), \mathbb{P}_\Pi(V^c)\} \to 0 \quad \text{as} \quad n \to \infty, \tag{222}$$

where $\mathbb{P}_\Pi$ is the product probability measure, i.e., $\mathbb{P}_\Pi\big[x_1, x_2, \cdots, x_n\big] = \prod_{i=1}^n \mathbb{P}[X_i = x_i]$ for all $(x_1, x_2, \cdots, x_n) \in \mathcal{X}^n$. Let $S_n = X_1 + X_2 + \cdots + X_n$ and

$$\tilde{T} = \frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}. \tag{223}$$

Then, under the condition that

$$\frac{1}{n^{3/2}}\sum_{i=1}^n \mathbb{E}[|X_i^3|] \to 0, \tag{224}$$

$$\frac{1}{n^2}\sum_{i=1}^n \mathbb{E}[|X_i|^4] \to 0, \tag{225}$$

we have

$$\tilde{T} \xrightarrow{\text{(d)}} \mathcal{N}(0,1). \tag{226}$$

This lemma can recover the original Stein's criterion [25, Theorem 3.2] for independent random variables by setting $\mathcal{V}^c = \emptyset$ and $\xi_n = 0$.

Proof: Appendix B.1. ■

Now, we return to proof Theorem 5. As in the i.i.d. case, Eq. (196) holds, where Z_{ij} in given in (198). Then,

$$\begin{aligned}Z_{ij} &= \sum_{k=1}^n \sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik}, X_{jk}) = (x,x')\} && (227)\\ &= n\sum_{Q_{XX'}}\sum_{x,x'} Q_{XX'}(x,x') d_{\mathrm{B}}(x,x') Z_{Q_{XX'}}, && (228)\end{aligned}$$

where $Z_{Q_{XX'}} = \mathbb{1}\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q_{XX'})\}$. Define

$$U_{Q_{XX'}} = \sqrt{\frac{|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|}{\sum_{Q_{XX'}} \mathbb{E}[V_{Q_{XX'}}^2]}}, \tag{229}$$

where

$$\begin{aligned}V_{Q_{XX'}} = &\sum_{x,x'} Q_{XX'}(x,x') d_{\mathrm{B}}(x,x') Z_{Q_{XX'}} \\ &- \sum_{x,x'} Q_{XX'}(x,x') d_{\mathrm{B}}(x,x') \mathbb{E}\big[Z_{Q_{XX'}}\big].\end{aligned} \tag{230}$$

Then, we have

$$\frac{Z_{ij} - \mathbb{E}[Z_{ij}]}{\sqrt{\mathrm{Var}(Z_{ij})}} = \frac{\sum_{Q_{XX'}} U_{Q_{XX'}}}{\sqrt{\mathrm{Var}\left(\sum_{Q_{XX'}} U_{Q_{XX'}}\right)}} \tag{231}$$

and also that $\mathbb{E}[U_{Q_{XX'}}] = 0$ and $\sum_{Q_{XX'}} \mathbb{E}[U_{Q_{XX'}}^2] = |\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|$. Now, define the set

$$\begin{aligned}\mathcal{V}_0 = \bigg\{&\{z_{Q_{XX'}}\}_{Q_{XX'}\in\mathcal{T}_n(\mathcal{X}\times\mathcal{X})} : \text{the only } n\text{-joint type } Q_{XX'} \\ &\text{such that } z^*_{Q_{XX'}} = 1 \text{ is } Q^*_{XX'} = Q_X Q_X\bigg\}.\end{aligned} \tag{232}$$

Then, for any $\{z_{Q_{XX'}}\}_{Q_{XX'}} \in \mathcal{V}_0$, we have that following probability, where $Q_{XX'} \in \mathcal{T}_n(\mathcal{X}\times\mathcal{X})$, satisifes

$$\begin{aligned}&\mathbb{P}\bigg[\bigcap_{Q_{XX'}} \{Z_{Q_{XX'}} = z_{Q_{XX'}}\}\bigg] \\ &= \mathbb{P}\bigg[\{Z_{Q^*_{XX'}} = 1\} \cap \bigcap_{Q_{XX'}\neq Q^*_{XX'}} \{Z_{Q_{XX'}} = 0\}\bigg] && (233)\\ &= \mathbb{P}\bigg[\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\} \\ &\qquad \times \cap \bigcap_{Q_{XX'}\neq Q^*_{XX'}} \{(\boldsymbol{X}_i, \boldsymbol{X}_j) \notin \mathcal{T}_n(Q_{XX'})\}\bigg] && (234)\\ &= \mathbb{P}\bigg[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\bigg]. && (235)\end{aligned}$$

Similarly, for any sequence $\{z_{Q_{XX'}}\}_{Q_{XX'}} \in \mathcal{V}_0$, we have

$$\begin{aligned}
&\prod_{Q_{XX'}} \mathbb{P}\Big[Z_{Q_{XX'}} = z_{Q_{XX'}}\Big] \\
&= \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\big] \\
&\quad \times \prod_{Q_{XX'} \neq Q^*_{XX'}} \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \notin \mathcal{T}_n(Q_{XX'})\big] && (236) \\
&= \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\big] \\
&\quad \times \prod_{Q_{XX'} \neq Q^*_{XX'}} \big(1 - 2^{-nI_{Q_{XX'}}(X;X')}\big) && (237) \\
&\geq \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\big]\big(1 - |\mathcal{P}_n(\mathcal{X} \times \mathcal{X})| 2^{-nI_{\min}}\big), && (238)
\end{aligned}$$

where (238) follows from the fact that $\prod_{i=1}^n (1 - a_i) \geq 1 - \sum_{i=1}^n a_i$ for any $a_1, a_2, \cdots, a_n \in [0,1]$. Here,

$$I_{\min} = \min_{Q_{XX'}: Q_{XX'} \neq Q^*_{XX'}} I_{Q_{XX'}}(X;X') > 0. \quad (239)$$

From (235) and (238), there is a positive sequence $\xi_n := |\mathcal{P}_n(\mathcal{X} \times \mathcal{X})| 2^{-nI_{\min}} \to 0$ as $n \to \infty$ such that

$$\begin{aligned}
&(1 - \xi_n) \prod_{Q_{XX'}} \mathbb{P}\big[Z_{Q_{XX'}} = z_{Q_{XX'}}\big] \\
&\leq \mathbb{P}\Big[\bigcap_{Q_{XX'}} \{Z_{Q_{XX'}} = z_{Q_{XX'}}\}\Big] && (240) \\
&\leq (1 + \xi_n) \prod_{Q_{XX'}} \mathbb{P}\big[Z_{Q_{XX'}} = z_{Q_{XX'}}\big]. && (241)
\end{aligned}$$

Furthermore, it follows from (235) and the definition of $\mathcal{V}_0$ in (232) that

$$\mathbb{P}[\{Z_{Q_{XX'}}\} \in \mathcal{V}_0] = \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\big]. \quad (242)$$

Therefore, we obtain

$$\begin{aligned}
\mathbb{P}[\{Z_{Q_{XX'}}\} \in \mathcal{V}_0^c] &= 1 - \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\big] && (243) \\
&= \frac{|\mathcal{T}_n(Q_X)|^2 - |\mathcal{T}_n(Q^*_{XX'})|}{|\mathcal{T}_n(Q_X)|^2}. && (244)
\end{aligned}$$

In addition, from (238), we obtain

$$\begin{aligned}
&\mathbb{P}_\Pi[\{Z_{Q_{XX'}}\} \in \mathcal{V}_0^c] \\
&= 1 - \mathbb{P}_\Pi[\{Z_{Q_{XX'}}\} \in \mathcal{V}_0] && (245) \\
&\leq 1 - \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(Q^*_{XX'})\big] \\
&\quad \times \big(1 - |\mathcal{P}_n(\mathcal{X} \times \mathcal{X})| 2^{-nI_{\min}}\big) && (246) \\
&= \frac{|\mathcal{T}_n(Q_X)|^2 - |\mathcal{T}_n(Q^*_{XX'})|}{|\mathcal{T}_n(Q_X)|^2} \\
&\quad + |\mathcal{P}_n(\mathcal{X} \times \mathcal{X})|^2 \frac{|\mathcal{T}_n(Q^*_{XX'})|}{|\mathcal{T}_n(Q_X)|^2} 2^{-nI_{\min}}. && (247)
\end{aligned}$$

Since $V_{Q_{XX'}}$ is linear in $Z_{Q_{XX'}}$ (cf. (230)), the existence of a set $\mathcal{V}$ as in Lemma 13 is guaranteed with the same $\mathbb{P}[\mathcal{V}^c] = \mathbb{P}[\{Z_{Q_{XX'}}\} \in \mathcal{V}_0^c]$ and $\mathbb{P}_\Pi[\mathcal{V}^c] = \mathbb{P}_\Pi[\{Z_{Q_{XX'}}\} \in \mathcal{V}_0^c]$.

Now, since $V_{Q_{XX'}}$ is bounded for all $Q_{XX'} \in \mathcal{T}_n(\mathcal{X} \times \mathcal{X})$. Hence, we have

$$\sum_{Q_{XX'}} \mathbb{E}[V^2_{Q_{XX'}}] = \Theta(|\mathcal{P}_n(\mathcal{X} \times \mathcal{X})|). \quad (248)$$

By the same fact, we also have

$$g_n = O\big(\big|\mathcal{T}_n(\mathcal{X} \times \mathcal{X})\big|\big). \quad (249)$$

Hence, it holds that

$$g_n \xi_n \to 0 \quad (250)$$

by the sub-exponential number of possible n-joint type.

From (244), (247), and (249), we obtain

$$\begin{aligned}
&g_n \max\{\mathbb{P}[V^c], \mathbb{P}_\Pi[V^c]\} \\
&= O\bigg(|\mathcal{P}_n(\mathcal{X} \times \mathcal{X})| \bigg(\frac{|\mathcal{T}_n(Q_X)|^2 - |\mathcal{T}_n(Q^*_{XX'})|}{|\mathcal{T}_n(Q_X)|^2} \\
&\quad + |\mathcal{P}_n(\mathcal{X} \times \mathcal{X})| \frac{|\mathcal{T}_n(Q^*_{XX'})|}{|\mathcal{T}_n(Q_X)|^2} 2^{-nI_{\min}}\bigg)\bigg) && (251) \\
&= O\bigg(|\mathcal{P}_n(\mathcal{X} \times \mathcal{X})| \bigg(\frac{|\mathcal{T}_n(Q_X)|^2 - |\mathcal{T}_n(Q^*_{XX'})|}{|\mathcal{T}_n(Q_X)|^2}\bigg)\bigg). && (252)
\end{aligned}$$

Under the regular condition of type in (19), it holds that

$$\frac{|\mathcal{T}_n(Q_X)|^2 - |\mathcal{T}_n(Q^*_{XX'})|}{|\mathcal{T}_n(Q_X)|^2} = \mathbb{P}[(\boldsymbol{X}_i, \boldsymbol{X}_j) \notin \mathcal{T}_n(Q^*_{XX'})] \to 0. \quad (253)$$

The regular condition of types assumes that the rate of convergence to zero of $\mathbb{P}[(\boldsymbol{X}_i, \boldsymbol{X}_j) \notin \mathcal{T}_n(Q^*_{XX'})]$ is faster than $O(1/(n+1)^{|\mathcal{X}|^2})$. As a result, as $n \to \infty$, $g_n = O\big(\big|\mathcal{T}_n(\mathcal{X} \times \mathcal{X})\big|\big) \to 0$.

On the other hand, let $d_{\max} = \max_{x,x'} d_{\mathrm{B}}(x,x')$. Then, we also have

$$\begin{aligned}
&\big|V_{Q_{XX'}}\big|^4 \\
&\leq 8\bigg(\bigg|\sum_{x,x'} Q_{XX'}(x,x') d_{\mathrm{B}}(x,x') Z_{Q_{XX'}}\bigg|^4 \\
&\quad + \bigg|\sum_{x,x'} Q_{XX'}(x,x') d_{\mathrm{B}}(x,x') \mathbb{P}\big(Z_{Q_{XX'}} = 1\big)\bigg|^4\bigg) && (254) \\
&\leq 16 d^4_{\max} && (255)
\end{aligned}$$

for all $Q_{XX'} \in \mathcal{T}_n(\mathcal{X} \times \mathcal{X})$, and

$$\begin{aligned}
&\big|V_{Q_{XX'}}\big|^3 \\
&\leq 4\bigg(\bigg|\sum_{x,x'} Q_{XX'}(x,x') d_{\mathrm{B}}(x,x') Z_{Q_{XX'}}\bigg|^3 \\
&\quad + \bigg|\sum_{x,x'} Q_{XX'}(x,x') d_{\mathrm{B}}(x,x') \mathbb{P}\big(Z_{Q_{XX'}} = 1\big)\bigg|^3\bigg) && (256) \\
&\leq 8 d^3_{\max}, && (257)
\end{aligned}$$

where (254) and (256) follow from Hölder inequality for counting measure [28, Sec. 7.2].

Hence, we have

$$\frac{1}{|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|^2}\sum_{Q_{XX'}}\mathbb{E}[U_{Q_{XX'}}^4]$$
$$=\left(\frac{|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|}{\sum_{Q_{XX'}}\mathbb{E}[V_{Q_{XX'}}^2]}\right)\frac{1}{|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|^2}\sum_{Q_{XX'}}\mathbb{E}[V_{Q_{XX'}}^4] \tag{258}$$
$$\dot{\leq}\frac{1}{|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|^2}16|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|d_{\max}^4 \tag{259}$$
$$\to 0 \tag{260}$$

as $n\to\infty$, where (259) follows from (255). Similarly, we have

$$\frac{1}{|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|^{3/2}}\sum_{Q_{XX'}}\mathbb{E}[|U_{Q_{XX'}}|^3]$$
$$\leq\frac{1}{|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|^{3/2}}8|\mathcal{P}_n(\mathcal{X}\times\mathcal{X})|d_{\max}^3 \tag{261}$$
$$\to 0 \tag{262}$$

as $n\to\infty$.

From the above facts and Lemma 13, we conclude that

$$T_{ij}=\frac{Z_{ij}-\mathbb{E}[Z_{ij}]}{\sqrt{\mathrm{Var}(Z_{ij})}}\xrightarrow{(\mathrm{d})}\mathcal{N}(0,1). \tag{263}$$

Similarly, we can prove that if M is a constant, we have

$$\sum_{i\neq j}\alpha_{ij}T_{ij}\xrightarrow{(\mathrm{d})}\mathcal{N}(0,1). \tag{264}$$

for any sequence $\{\alpha_{ij}\}_{i,j\in[M],i\neq j}$. Using the same arguments as the proof of Lemma 5 , we obtain (20).

F. Proof of Theorem 6

Our proof of this theorem is based on a modification of the Wasserstein metric, inspired by the classical Kolmogorov and Wasserstein metrics, that measures the distance between the distribution of the error exponent and that of the standard Gaussian. Such modification is needed to deal with an infinite number of terms as $n\to\infty$, a case where the classical Wasserstein metric upper bound fails to work [25, Prop. 2.4]. After introducing important lemmas from (265) to (270), we start our proof in (271) to obtain (283) and (289). The asymptotics of the random variables $T_{ij}(n)$ in (283) and (289) are studied in four steps: the first step is split into two sub-steps in (290)–(308) and (309)–(327), the second step from (328) to (343), the third step from (344) to (350) and the last step to obtain (22) in (351)–(353).

Recall the definitions of probability metrics in Definition 1. First, we prove the following fundamental lemma.

Lemma 14: If $Z\sim\mathcal{N}(0,1)$, then for any random variable T, it holds that

$$\big|\mathbb{P}(T\leq x)-\mathbb{P}(Z\leq x)\big|\leq 2(8\pi)^{-1/4}\sqrt{d_{W,\mathrm{mod}}(T,Z)}+\big|\mathbb{P}(T\leq x)-\mathbb{P}(T\geq -x)\big| \tag{265}$$

for all $x\in\mathbb{R}$. In addition, if the distribution of T is tight[4], for any $x\to 0$, which is a continuous point of the limit distribution of T, as $n\to\infty$, we have

$$\limsup_{n\to\infty}\big|\mathbb{P}(T\leq x)-\mathbb{P}(Z\leq x)\big|\leq 2(8\pi)^{-1/4}\limsup_{n\to\infty}\sqrt{d_{W,\mathrm{mod}}(T,Z)}. \tag{266}$$

Proof: Appendix B.2. ■

By using the definition of $d_{W,\mathrm{mod}}$ and setting $T=X$, we obtain the following result, which is tighter than (or at least equal to) the upper bound of $d_K(T,Z)$ in [25, Prop. 2.4]. However, we note that the probability metric here is the modified Wasserstein metric. See the same arguments to achieve a similar result in [25, Prop. 2.4].

Lemma 15: For $h\in\mathcal{H}$, let f_h solve

$$f_h'(w)-wf_h(w)=h(w)-\mathbb{E}[h(Z)]. \tag{267}$$

If T is a random variable and Z has the standard normal distribution, then

$$d_{W,\mathrm{mod}}(T,Z)=\sup_{h\in\mathcal{H}}\min\Big\{\big|\mathbb{E}\big[f_h'(T)-Tf_h(T)\big]\big|,\big|\mathbb{E}\big[f_h'(-T)+Tf_h(-T)\big]\big|\Big\}. \tag{268}$$

Proof: Appendix B.3. ■

Now, we prove the following lemma.

Lemma 16: Assume that $T=\min\{T_1,T_2,\cdots,T_L\}$ for some $L\in\mathbb{Z}^+$ and $T_1,T_2,\cdots,T_L$ are identically distributed random variables. Then, it holds that

$$d_{W,\mathrm{mod}}(T,Z)\leq\max\Big\{\sup_{h\in\mathcal{H}}\big|\mathbb{E}[f_h'(T_1)-T_1f_h(T_1)]\big|,\sup_{h\in\mathcal{H}}\big|\mathbb{E}[f_h'(-T_1)+T_1f_h(-T_1)]\big|\Big\}+\sup_{h\in\mathcal{H}}\min\Big\{\mathbb{E}[h(T)-h(T_1)],\mathbb{E}[h(-T_1)-h(-T)]\Big\}. \tag{269}$$

Proof: Appendix B.4. ■

Lemma 17: [25, Th. 3.2] Let $X_1,X_2,\cdots,X_n$ be independent mean zero random variables such that $\mathbb{E}[|X_i|^4]<\infty$ and $\mathbb{E}[X_i^2]=1$. If $T=\sum_{i=1}^n X_i/\sqrt{n}$ and Z has the standard normal distribution, then

$$\max\Big\{\sup_{h\in\mathcal{H}}\big|\mathbb{E}\big[f_h'(T)-Tf_h(T)\big]\big|,\sup_{h\in\mathcal{H}}\big|\mathbb{E}\big[f_h'(-T)+Tf_h(-T)\big]\big|\Big\}\leq\frac{1}{n^{3/2}}\sum_{i=1}^n\mathbb{E}[|X_i|^3]+\frac{\sqrt{2}}{n\sqrt{\pi}}\sqrt{\sum_{i=1}^n\mathbb{E}[X_i^4]}. \tag{270}$$

[4]A distribution on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ is tight if for any fixed $\varepsilon>0$, there exists $u,v\in\mathbb{R}$ such that $\mathbb{P}(u<T\leq v)>1-\varepsilon$ [24].

We can observe the fact (270) since T and $-T$ are both the sums of independent random variables. Now, we are ready to prove Theorem 6. Observe that

$$\max_{i\neq j}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)\le P_{\mathrm{e}}(\mathcal{C}_n)$$
$$\le\sum_{i=1}^{M_n}\sum_{j\neq i}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j)$$
$$\le M_n(M_n-1)\max_{i\neq j}\mathbb{P}(\boldsymbol{X}_i\to\boldsymbol{X}_j). \tag{271}$$

For M_n sub-exponential in n, Eq. (196) in Theorem 5 used for a constant number of messages, is still a valid result here for M_n sub-exponential in n. We explicitly state the dependence on n in Z_{ij} in (198) as $Z_{ij}(n)$. We define

$$T_{ij}(n)=\frac{Z_{ij}(n)-\mathbb{E}[Z_{ij}(n)]}{\sqrt{\mathrm{Var}(Z_{ij}(n))}}, \tag{272}$$

we have

$$\min_{i\neq j}T_{ij}(n)=\min_{i\neq j}\frac{Z_{ij}(n)-\mathbb{E}[Z_{ij}(n)]}{\sqrt{\mathrm{Var}(Z_{ij}(n))}}. \tag{273}$$

Now, for any $\varepsilon>0$, let the event

$$\mathcal{E}_n=\left\{\frac{1}{M_n(M_n-1)}\left|\sum_{i\neq j}T_{ij}(n)\right|\ge\varepsilon\right\} \tag{274}$$

for all $n\in\mathbb{Z}^+$. Then, we have

$$\sum_{n=1}^{\infty}\mathbb{P}(\mathcal{E}_n)\le\sum_{n=1}^{\infty}\mathbb{P}\left[\frac{1}{(M_n-1)M_n}\left|\sum_{i\neq j}T_{ij}(n)\right|\ge\varepsilon\right] \tag{275}$$
$$\le\sum_{n=1}^{\infty}\frac{1}{\varepsilon^2M_n^2(M_n-1)^2}\mathrm{Var}\left(\sum_{i\neq j}T_{ij}(n)\right) \tag{276}$$
$$=\sum_{n=1}^{\infty}\frac{1}{\varepsilon^2M_n^2(M_n-1)^2}\sum_{i\neq j}\mathrm{Var}\big(T_{ij}(n)\big) \tag{277}$$
$$=\sum_{n=1}^{\infty}\frac{1}{\varepsilon^2M_n(M_n-1)} \tag{278}$$
$$<\infty, \tag{279}$$

where (276) follows from Chebyshev's inequality, (277) follows from the pairwise independence of Z_{ij}'s, and (279) follows from the condition (21). Hence, by the Borel–Cantelli lemma, from (279), we have

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\mathcal{E}_k^c\right]=1. \tag{280}$$

However, we have

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\mathcal{E}_k^c\right]$$
$$=\mathbb{P}\left[\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\frac{1}{M_n(M_n-1)}\left|\sum_{i\neq j}T_{ij}(k)\right|<\varepsilon\right\}\right]. \tag{281}$$

It follows from (280) and (281) that

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\frac{1}{M_n(M_n-1)}\left|\sum_{i\neq j}T_{ij}(k)\right|<\varepsilon\right\}\right]=1, \tag{282}$$

or

$$\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\xrightarrow{\text{(a.s.)}}0, \tag{283}$$

as $n\to\infty$. Now, from Theorem 5, we have $T_{ij}(n),T_{i'j'}(n),T_{12}(n)$ are independent as $n\to\infty$ if $(i,j)\neq(i',j')\neq(1,2)$. Then, for any $B_1,B_2\in\mathcal{B}(R)$ (Borel sets in $\mathbb{R}$), as $n\to\infty$, we have

$$\mathbb{P}\left[\Big\{T_{ij}(n)-T_{12}(n)\in B_1\Big\}\cap\Big\{T_{i'j'}(n)-T_{12}(n)\in B_2\Big\}\right]$$
$$=\int_{\mathbb{R}}\mathbb{P}\left[\Big\{T_{ij}(n)-T_{12}(n)\in B_1\Big\}\cap\Big\{T_{i'j'}(n)-T_{12}(n)\in B_2\Big\}\middle|T_{12}(n)=\alpha\right]f_{T_{12}(n)}(\alpha)d\alpha \tag{284}$$
$$=\int_{\mathbb{R}}\mathbb{P}\left[\Big\{T_{ij}(n)\in\alpha+B_1\Big\}\cap\Big\{T_{i'j'}(n)\in\alpha+B_2\Big\}\middle|T_{12}(n)=\alpha\right]f_{T_{12}(n)}(\alpha)d\alpha \tag{285}$$
$$=\int_{\mathbb{R}}\mathbb{P}\left[\Big\{T_{ij}(n)\in\alpha+B_1\Big\}\cap\Big\{T_{i'j'}(n)\in\alpha+B_2\Big\}\right]f_{T_{12}(n)}(\alpha)d\alpha \tag{286}$$
$$=\int_{\mathbb{R}}\mathbb{P}\big[T_{ij}(n)\in\alpha+B_1\big]\mathbb{P}\big[T_{i'j'}(n)\in\alpha+B_2\big]f_{T_{12}}(\alpha)d\alpha \tag{287}$$
$$=\mathbb{P}\big[T_{ij}(n)-T_{12}(n)\in B_1\big]\times\mathbb{P}\big[T_{i'j'}(n)-T_{12}(n)\in B_2\big]+o(1), \tag{288}$$

i.e., $T_{ij}(n)-T_{12}(n)$ and $T_{i'j'}(n)-T_{12}(n)$ are asymptotically independent. This means that $\{T_{ij}(n)-T_{12}(n)\}$ are asymptotically pairwise independent. Hence, by using the same arguments to achieve (283), we have

$$\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)-T_{12}(n)\xrightarrow{\text{(a.s.)}}0, \tag{289}$$

as $n\to\infty$ (point-wise convergence).

The first step consists of showing that $\max\big\{\mathbb{E}[h(\min_{i\neq j}T_{ij}(n))-h(T_{12}(n))],\mathbb{E}[h(-T_{12}(n))-h(-\min_{i\neq j}T_{ij}(n))]\big\}\to0$ as $n\to\infty$. Since $h(u)_{-a,-c}=1-h(-u)_{a,c}$ where $h(u)_{a,c}$ is the h function in $\mathcal{V}$ with parameter a,c defined in (35), it is enough to carry out with two sub-steps, step 1a and step 1b.

1) Step 1a: To begin with, we prove that $\mathbb{E}[h(\min_{i\neq j}T_{ij}(n))-h(T_{12}(n))]\to0$ as $n\to\infty$ for all $h\in\{\mathcal{H}:a\ge0\}$. We have two different cases based on the value of a: that $\liminf_{n\to\infty}a>0$ and that $\lim_{n\to\infty}a=0$.

For the first case, from (283), we have $\min_{i\neq j}T_{ij}(n)<a$ as $n\to\infty$. It follows that

$$h(\min_{i\neq j}T_{ij}(n))=h\bigg(\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg)=c \tag{290}$$

by the definition of $\mathcal{H}$. Then, we have

$$h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))$$
$$= \left[h(\min_{i\neq j} T_{ij}(n)) - h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right)\right] + \left[h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) - h(T_{12}(n))\right] \tag{291}$$
$$= h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) - h(T_{12}(n)) \tag{292}$$
$$\leq \left|h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) - h(T_{12}(n))\right|. \tag{293}$$

For the second case, if $\min_{i\neq j} T_{ij}(n) \leq a$ as $n \to \infty$, we have

$$h(\min_{i\neq j} T_{ij}(n)) - h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right)$$
$$= h(a) - h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) \tag{294}$$
$$\leq \left|a - \frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right| \tag{295}$$
$$\leq \max\left\{a, \frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right\} \tag{296}$$
$$\to 0 \tag{297}$$

as $n \to \infty$, where (295) follows from 1-Lipschitz property of h for all $h \in \mathcal{V}$. On the other hand, if $a < \min_{i\neq j} T_{ij}(n) \leq \frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)$ and $\liminf_{n\to\infty} c > 0$, we have

$$h(\min_{i\neq j} T_{ij}(n)) - h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right)$$
$$= \frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n) - \min_{i\neq j} T_{ij}(n) \tag{298}$$
$$\to 0 \tag{299}$$

as $n \to \infty$.

In addition, if $a < \min_{i\neq j} T_{ij}(n) \leq \frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)$ and $\lim_{n\to\infty} c = 0$, we have

$$h(\min_{i\neq j} T_{ij}(n)) - h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) \leq c \tag{300}$$
$$\to 0 \tag{301}$$

as $n \to \infty$.

From (297), (299), and (301), it holds that

$$\limsup_{n\to\infty} h(\min_{i\neq j} T_{ij}(n)) - h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) \leq 0. \tag{302}$$

Combining (293) and (302), we obtain

$$\limsup_{n\to\infty} h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))$$
$$\leq \limsup_{n\to\infty} \left|h\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) - h(T_{12}(n))\right| \tag{303}$$
$$\leq \limsup_{n\to\infty} \left|\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n) - T_{12}(n)\right| \tag{304}$$
$$= 0, \tag{305}$$

where (305) follows from (289).

Now, since $|\big(h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))\big)| \leq c \leq 4\sqrt{2\pi}$ for all $h \in \mathcal{V}$, hence by the reverse Fatou's lemma [24, Theorem 5.4], we have

$$\limsup_{n\to\infty} \mathbb{E}\left[h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))\right]$$
$$\leq \mathbb{E}\left[\limsup_{n\to\infty} h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))\right] \tag{306}$$
$$= 0, \tag{307}$$

where (307) follows from (305). Since $h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n)) \geq 0$, by the fact that h is non-increasing for all $h \in \mathcal{V}$, from (307), we obtain

$$\lim_{n\to\infty} \mathbb{E}\big[h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))\big] = 0. \tag{308}$$

2) Step 1b: Next, we prove that $\mathbb{E}[h(-\min_{i\neq j} T_{ij}(n)) - h(-T_{12}(n))] \to 0$ as $n \to \infty$ for all $h \in \{\mathcal{H} : a < 0\}$.

For all $h \in \mathcal{H}$, let $\tilde{h}(x) = h(-x)$ for all $x \in \mathbb{R}$. Then, we have

$$h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n)) = \tilde{h}(T_{12}(n)) - \tilde{h}(\min_{i\neq j}\{T_{ij}\}(n)). \tag{309}$$

Now, we show that $\mathbb{E}[\tilde{h}(T_{12}(n)) - \tilde{h}(\min_{i\neq j} T_{ij}(n))] \to 0$ as $n \to \infty$. Similar to Step 1a, we divide into different cases based on the value of $a + c$.

For $\limsup_{n\to\infty}(a + c) < 0$, from (283), as $n \to \infty$, we have

$$\min_{i\neq j} T_{ij}(n) \leq \frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n) < -(a+c). \tag{310}$$

Hence, it holds that

$$\tilde{h}\big(\max_{i\neq j} \tilde{T}_{ij}(n)\big) = \tilde{h}\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) = 0. \tag{311}$$

It follows that as $n \to \infty$, we have

$$\tilde{h}\left(\frac{1}{M_n(M_n-1)}\sum_{i\neq j} T_{ij}(n)\right) - \tilde{h}(\min_{i\neq j} T_{ij}(n)) = 0. \tag{312}$$

For the second case where $\lim_{n\to\infty} a+c = 0$, if $\min_{i\neq j} T_{ij}(n) \leq -(a+c)$, as $n\to\infty$, we have

$$\tilde{h}\bigg(\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg)-\tilde{h}(\min_{i\neq j}\{T_{ij}(n)\})$$
$$=\tilde{h}\bigg(\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg)-\tilde{h}(-(a+c)) \tag{313}$$
$$\leq\bigg|\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)+(a+c)\bigg| \tag{314}$$
$$\leq\bigg|\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg|+|a+c| \tag{315}$$
$$\to 0 \tag{316}$$

as $n\to\infty$.

In addition, if $\min_{i\neq j} T_{ij}(n)\geq -(a+c)$, we have

$$\tilde{h}\bigg(\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg)-\tilde{h}(\min_{i\neq j}T_{ij}(n))$$
$$\leq\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)-\min_{i\neq j}T_{ij}(n) \tag{317}$$
$$\leq\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)+(a+c) \tag{318}$$
$$\to 0 \tag{319}$$

as $n\to\infty$.

From (316) and (319), as $n\to\infty$, we have

$$\limsup_{n\to\infty}\tilde{h}\bigg(\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg)-\tilde{h}(\min_{i\neq j}\{T_{ij}(n)\})\leq 0. \tag{320}$$

It follows from (312) and (320) that

$$\limsup_{n\to\infty}\tilde{h}(T_{12}(n))-\tilde{h}(\min_{i\neq j}T_{ij}(n))$$
$$\leq\limsup_{n\to\infty}\bigg[\tilde{h}(T_{12}(n))-\tilde{h}\bigg(\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg)\bigg]$$
$$+\limsup_{n\to\infty}\bigg[\tilde{h}\bigg(\frac{1}{M_n(M_n-1)}\sum_{i\neq j}T_{ij}(n)\bigg)-\tilde{h}(\min_{i\neq j}T_{ij}(n))\bigg] \tag{321}$$
$$=0, \tag{322}$$

where (322) follows from (320) and (289).

Now, since $|\tilde{h}(\min_{i\neq j}T_{ij}(n))-\tilde{h}(T_{12}(n))|\leq c\leq 4\sqrt{2\pi}$ for all $h\in\mathcal{V}$, hence by the reverse Fatou's lemma [24, Theorem 5.4], we have

$$\limsup_{n\to\infty}\mathbb{E}\Big[\tilde{h}(T_{12}(n))-\tilde{h}(\min_{i\neq j}T_{ij}(n))-\Big]$$
$$\leq\mathbb{E}\Big[\limsup_{n\to\infty}\tilde{h}(T_{12}(n))-\tilde{h}(\min_{i\neq j}T_{ij}(n))\Big] \tag{323}$$
$$=0, \tag{324}$$

where (324) follows from (322). Since $h(\min_{i\neq j}T_{ij}(n))-h(T_{12}(n))\geq 0$, from (324), we obtain

$$\lim_{n\to\infty}\mathbb{E}\big[\tilde{h}(T_{12}(n))-\tilde{h}(\min_{i\neq j}T_{ij}(n))\big]=0, \tag{325}$$

or

$$\lim_{n\to\infty}\mathbb{E}\big[h(-T_{12}(n))-h(-\min_{i\neq j}T_{ij}(n))\big]=0. \tag{326}$$

From (308) and (326), we finally have that, for all $h\in\mathcal{H}$:

$$\lim_{n\to\infty}\min\bigg\{\mathbb{E}\big[h(\min_{i\neq j}T_{ij}(n))-h(T_{12}(n))\big],\ \mathbb{E}\big[h(-T_{12}(n))-h(-\min_{i\neq j}T_{ij}(n))\big]\bigg\}=0. \tag{327}$$

3) Step 2: In this step, we show that $\lim_{n\to\infty} d_{W,\text{mod}}(\min_{i\neq j}T_{ij},Z)=0$. Indeed, from Lemma 17, we have

$$\sup_{h\in\mathcal{H}}\big|\mathbb{E}\big[f_h'(T_{ij}(n))-T_{ij}(n)f_h(T_{ij}(n))\big]\big|$$
$$\leq\frac{1}{n^{3/2}}\sum_{k=1}^n\mathbb{E}[|X_k|^3]+\frac{\sqrt{2}}{n\sqrt{\pi}}\sqrt{\sum_{k=1}^n\mathbb{E}[X_k^4]} \tag{328}$$
$$=\frac{1}{\sqrt{n}}\mathbb{E}[|X_1|^3]+\frac{\sqrt{2}}{\sqrt{\pi n}}\sqrt{\mathbb{E}[X_1^4]} \tag{329}$$

where

$$X_k:=\frac{\sum_{x,x'}d_{\mathrm{B}}(x,x')\big(\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\big)}{\sqrt{\mathrm{Var}\big(-\sum_{x,x'}d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\big)}}-\frac{\sum_{x,x'}d_{\mathrm{B}}(x,x')\mathbb{P}\big[(X_{ik},X_{jk})=(x,x')\big]\big)}{\sqrt{\mathrm{Var}\big(-\sum_{x,x'}d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\big)}} \tag{330}$$

for all $k\in[n]$.

Now, observe that

$$\mathrm{Var}(Z_{ij}(n))$$
$$=\mathrm{Var}\bigg(\sum_{k=1}^n\sum_{x,x'}d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\bigg) \tag{331}$$
$$=\sum_{k=1}^n\mathrm{Var}\bigg(\sum_{x,x'}d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\bigg) \tag{332}$$
$$=n\,\mathrm{Var}\bigg(\sum_{x,x'}d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\bigg),\quad\forall k, \tag{333}$$

where (332) and (333) follow from the fact that (X_{ik},X_{jk}) are i.i.d. given i,j.

Now, recall the definition of K_1 and K_2 in (152) and (153), respectively. Then, we have

$$\mathrm{Var}\bigg(\sum_{x,x'}d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\bigg)=K_1-K_2^2=L_2. \tag{334}$$

In addition, we have

$$\mathbb{E}\Big[\Big|\sum_{x,x'} d_{\mathrm{B}}(x,x')\big(\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\} - \mathbb{P}\big[(X_{ik},X_{jk})=(x,x')\big]\big)\Big|^3\Big] \tag{335}$$

$$\leq 4\bigg(\mathbb{E}\Big[\Big|\sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\}\Big|^3\Big] + \Big|\sum_{x,x'} d_{\mathrm{B}}(x,x')\mathbb{P}\big[(X_{ik},X_{jk})=(x,x')\big]\Big|^3\bigg) \tag{336}$$

$$= 4\bigg[\sum_{x,x'} d_{\mathrm{B}}^3(x,x')Q(x)Q(x') + \bigg(\sum_{x,x'} d_{\mathrm{B}}(x,x')Q(x)Q(x')\bigg)^3\bigg] = L_3, \tag{337}$$

where (336) follows from $(a+b)^3 \leq 4(|a|^3+|b^3|)$. Similarly, we have

$$\mathbb{E}\Big[\Big|\sum_{x,x'} d_{\mathrm{B}}(x,x')\big(\mathbb{1}\{(X_{ik},X_{jk})=(x,x')\} - \mathbb{P}\big[(X_{ik},X_{jk})=(x,x')\big]\big)\Big|^4\Big] \tag{338}$$

$$\leq 8\bigg[\sum_{x,x'} d_{\mathrm{B}}^4(x,x')Q(x)Q(x') + \bigg(\sum_{x,x'} d_{\mathrm{B}}(x,x')Q(x)Q(x')\bigg)^4\bigg] = L_4, \tag{339}$$

where we use $(a+b)^4 \leq 8(a^4+b^4)$ in (339).

Hence, from (329), (334), (337), and (339), we obtain

$$\sup_{h\in\mathcal{H}} \big|\mathbb{E}\big[f_h'(T_{ij}(n)) - T_{ij}(n)f_h(T_{ij}(n))\big]\big| \leq \frac{1}{\sqrt{n}}\bigg(\frac{L_3}{L_2^{3/2}}\bigg) + \sqrt{\frac{2}{\pi n}}\frac{L_4}{L_2^2}, \qquad \forall i\neq j. \tag{340}$$

Similarly, we also have

$$\sup_{h\in\mathcal{H}} \big|\mathbb{E}\big[f_h'(-T_{ij}(n)) + T_{ij}(n)f_h(-T_{ij}(n))\big]\big| \leq \frac{1}{\sqrt{n}}\bigg(\frac{L_3}{L_2^{3/2}}\bigg) + \sqrt{\frac{2}{\pi n}}\frac{L_4}{L_2^2}, \qquad \forall i\neq j. \tag{341}$$

Since $T_{ij}(n)$'s (for $i\neq j$) are identically distributed by the random codebook generation, it follows from Lemma 16 and (340) that for any $x\in\mathbb{R}$,

$$\begin{aligned} d_{W,\mathrm{mod}}(\min_{i\neq j} T_{ij}, Z) &\leq \max\bigg\{\sup_{h\in\mathcal{H}} \big|\mathbb{E}\big[f_h'(T_{12}(n)) - T_{12}(n)f_h(T_{12}(n))\big]\big|, \sup_{h\in\mathcal{H}} \big|\mathbb{E}\big[f_h'(-T_{12}(n)) + T_{12}(n)f_h(-T_{12}(n))\big]\big|\bigg\} \\ &\quad + \sup_{h\in\mathcal{H}} \min\bigg\{\mathbb{E}[h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))], \mathbb{E}[h(-T_{12}(n)) - h(-\min_{i\neq j} T_{ij}(n))]\bigg\} \end{aligned} \tag{342}$$

$$\leq \frac{1}{\sqrt{n}}\bigg(\frac{L_3}{L_2^{3/2}}\bigg) + \sqrt{\frac{2}{\pi n}}\frac{L_4}{L_2^2} + \sup_{h\in\mathcal{H}} \min\bigg\{\mathbb{E}[h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))], \mathbb{E}[h(-T_{12}(n)) - h(-\min_{i\neq j} T_{ij}(n))]\bigg\} \to 0, \tag{343}$$

where (343) follows from (327).

4) Step 3: In the third step, we prove that $\lim_{n\to\infty} \big|\mathbb{P}(\min_{i\neq j} T_{ij}(n) \leq x) - \mathbb{P}(\min_{i\neq j} T_{ij}(n) \geq -x)\big| = 0$ for all $x\in\mathbb{R}$ and x is a continuous point of the limiting distribution of $\min_{i\neq j} T_{ij}(n)$.

By the first step, we know that $\max\big\{\mathbb{E}[h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))], \mathbb{E}[h(-T_{12}(n)) - h(-\min_{i\neq j} T_{ij}(n))]\big\} \to 0$ as $n\to\infty$ for any $h\in\mathcal{V}$. Hence, by the proof of [25, Prop. 1.2], we have

$$\big|\mathbb{P}\big(\min_{i\neq j} T_{ij}(n) \leq x\big) - \mathbb{P}(T_{12}(n)\leq x)\big| \leq \frac{1}{\varepsilon}\sup_{h\in\mathcal{V}} \big|\mathbb{E}[h(\min_{i\neq j} T_{ij}(n)) - h(T_{12}(n))]\big| + O(\varepsilon), \tag{344}$$

$$\big|\mathbb{P}\big(\min_{i\neq j} T_{ij}(n) \geq -x\big) - \mathbb{P}(T_{12}(n)\leq -x)\big| \leq \frac{1}{\varepsilon}\sup_{h\in\mathcal{V}} \big|\mathbb{E}[-h(T_{12}(n)) - h(-\min_{i\neq j} T_{ij}(n))]\big| + O(\varepsilon). \tag{345}$$

Since ε is arbitrary chosen and the above limit fact, from (344) and (345), we obtain

$$\mathbb{P}\big(\min_{i\neq j} T_{ij}(n) \leq x\big) - \mathbb{P}\big(T_{12}(n)\leq x\big) \to 0, \tag{346}$$

$$\mathbb{P}\big(\min_{i\neq j} T_{ij}(n) \geq -x\big) - \mathbb{P}\big(T_{12}(n)\leq -x\big) \to 0. \tag{347}$$

From (346) and (347), we obtain

$$\lim_{n\to\infty} \mathbb{P}(\min_{i\neq j} T_{ij}(n) \leq x) - \mathbb{P}(\min_{i\neq j} T_{ij}(n) \geq -x) = \lim_{n\to\infty} \mathbb{P}(T_{12}(n)\leq x) - \mathbb{P}(T_{12}(n)\geq -x) \tag{348}$$

$$= Q(x) - Q(x) \tag{349}$$

$$= 0, \tag{350}$$

where (349) follows from CLT. Note that the form of T_{12} is defined in (272) is the normalized sum of i.i.d. random variables.

5) Step 4: The last step proves that $T_n = \frac{V_n - \mathbb{E}[V_n]}{\sqrt{\mathrm{Var}(V_n)}} \xrightarrow{\mathrm{(d)}} \mathcal{N}(0,1)$. From (343), Lemma 14, and Step 3, we have

$$\mathbb{P}\big[\min_{i\neq j} T_{ij}(n) \leq x\big] - \mathbb{P}\big[Z \leq x\big] \to 0 \tag{351}$$

as $n \to \infty$ for any continuous point $x \in \mathbb{R}$ of the limiting distribution of $\min_{i\neq j} T_{ij}(n)$, namely

$$\min_{i\neq j} T_{ij}(n) \xrightarrow{\mathrm{(d)}} Z = \mathcal{N}(0,1). \tag{352}$$

Using Lemma 1 and the same arguments to achieve (219) from (352) in the proof of Theorem 5, we obtain

$$T_n \xrightarrow{\mathrm{(d)}} \mathcal{N}(0,1). \tag{353}$$

Finally, from (196) and (353), by applying Slutsky's theorem [24, p. 334], we obtain (22).

G. Proof of Theorem 7

Consider first the case $0 \leq R < C$. Since the random variable $P_{\mathrm{e}}(\mathcal{C}_n)$ takes values in $[0,1]$, we have that

$$\begin{aligned} \mathrm{Var}[P_{\mathrm{e}}(\mathcal{C}_n)] &= \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^2] - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^2 && (354)\\ &\leq \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^2] && (355)\\ &\leq \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] \to 0 && (356)\end{aligned}$$

where (356) follows from the assumption that $\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] \to 0$ for $0 \leq R < C$. Applying Chebyshev's inequality we have that

$$\begin{aligned} \mathbb{P}\Big[|P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]| \geq \delta\Big] &\leq \frac{\mathrm{Var}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\delta^2} && (357)\\ &\leq \frac{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\delta^2} \to 0 && (358)\end{aligned}$$

where (358) follows from (356) and is valid for any given $\delta > 0$.

Now let us consider the case $R > C$. The following hold:

$$\begin{aligned} \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] &\to 1 && (359)\\ \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^2 &\to 1 && (360)\\ \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^2] &\geq \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^2 && (361)\\ \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^2] &\to 1 && (362)\\ \mathrm{Var}[P_{\mathrm{e}}(\mathcal{C}_n)] &\to 0 && (363)\end{aligned}$$

where (359) follows from the theorem assumption, (360) follows from (359), (361) follows from Jensen's inequality, (362) follows from (360) and (361) and the fact that $\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^2] \leq 1$, while (363) follows from (362) and (360) and the additivity of limits. Finally, using Chebyshev's inequality again we find that, for any $\delta > 0$,

$$\mathbb{P}\Big[|P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]| \geq \delta\Big] \leq \frac{\mathrm{Var}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\delta^2} \to 0. \tag{364}$$

H. Proof of Theorem 9

From [15, Th. 1] and from (91), for n sufficiently large we have:

$$\begin{aligned} &\mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq \gamma_n^\rho \min_{\rho\in[1,\infty)} E[P_e(\mathcal{C}_n)^{\frac{1}{\rho}}]^\rho\right]\\ &\quad= \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq 2^{-n(E_{\mathrm{trc}}(R)-\epsilon_n)}\right] && (365)\\ &\quad\leq \frac{1}{\gamma_n} && (366)\end{aligned}$$

where $\gamma_n \to \infty$, $\frac{\log \gamma_n}{n} \to 0$ and $\epsilon_n \to 0$. The Paley-Zygmund inequality [23, p. 1] implies that, for large enough n:

$$\begin{aligned} &\mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq \delta_n \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right]\\ &\quad= \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq 2^{-n(E_{\mathrm{rce}}(R)+\epsilon_n')}\right] && (367)\\ &\quad\geq (1-\delta_n)^2 \frac{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^2}{E[P_{\mathrm{e}}(\mathcal{C}_n)^2]} && (368)\end{aligned}$$

where we choose a sequence δ_n that goes to zero subexponentially, i.e., $\epsilon_n' \to 0$ and $0 < \delta_n < 1$ $\forall n$. Let n_0 be such that $\Delta E > \epsilon_n$, $\forall n > n_0$. Note that such an n_0 must exist from the definition of limit for ϵ_n. Now consider the following chain of inequalities for a large enough n, $n > n_0$:

$$\begin{aligned} 2^{-n(E_{\mathrm{trc}}(R)-\epsilon_n)} &= 2^{-n(E_{\mathrm{rce}}(R)+\Delta E-\epsilon_n)} && (369)\\ &\leq 2^{-n(E_{\mathrm{rce}}(R)+\Delta E-\epsilon_{n_0})} && (370)\\ &< 2^{-n(E_{\mathrm{rce}}(R)+\epsilon_n')} && (371)\end{aligned}$$

where (369) is from the theorem statement, (370) is valid from a certain n onwards from the definition of limit for ϵ_n, while (371) is because $\Delta E - \epsilon_{n_0}$ is a positive constant and, for large enough n, $\epsilon_n' < \Delta E - \epsilon_{n_0}$. Now, using (371), (365) and (367) we have:

$$\begin{aligned} &(1-\delta_n)\frac{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^2}{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^2]}\\ &\quad\leq \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq \delta_n \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right] && (372)\\ &\quad= \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq 2^{-n(E_{\mathrm{rce}}(R)+\epsilon_n')}\right] && (373)\\ &\quad\leq \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq 2^{-n(E_{\mathrm{trc}}(R)-\epsilon_n)}\right] && (374)\\ &\quad= \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) \geq \gamma_n^\rho \min_{\rho\in[1,\infty)} \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^{\frac{1}{\rho}}]^\rho\right] && (375)\\ &\quad\leq \frac{1}{\gamma_n} && (376)\end{aligned}$$

where(374) follows from (371). Finally, notice that, by definition, $(1-\delta_n) \to 1$ and $\frac{1}{\gamma_n} \to 0$ that imply:

$$\frac{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^2}{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^2]} \to 0.$$

I. Proof of Theorem 8

First, by the condition (24), we observe that

$$\begin{aligned} \frac{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}{\big(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\big)^2} &= \frac{\mathbb{E}[P_{\mathrm{e}}^2(\mathcal{C}_n)] - \big(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\big)^2}{\big(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\big)^2} && (377)\\ &= \frac{\mathbb{E}[P_{\mathrm{e}}^2(\mathcal{C}_n)]}{\big(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\big)^2} - 1 && (378)\\ &\to 0. && (379)\end{aligned}$$

On the other hand, by Theorem 9 we know that $\frac{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}{\left(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right)^2} \to \infty$ if $E_{\mathrm{rce}}(R) < E_{\mathrm{trc}}(R)$. Hence, from (379), we must have

$$E_{\mathrm{trc}}(R) = E_{\mathrm{rce}}(R). \tag{380}$$

Now, for any $\varepsilon > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left[\left| -\frac{\log P_{\mathrm{e}}(\mathcal{C}_n)}{n} - E_{\mathrm{trc}}(R)\right| > \varepsilon\right] \\
&\quad= \mathbb{P}\left[\left\{P_{\mathrm{e}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right\} \cup \left\{P_{\mathrm{e}}(\mathcal{C}_n) > 2^{-n(E_{\mathrm{trc}}(R)-\varepsilon)}\right\}\right] && (381)\\
&\quad= \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right] + \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) > 2^{-n(E_{\mathrm{trc}}(R)-\varepsilon)}\right] && (382)\\
&\quad\le \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right] + 2^{n(E_{\mathrm{trc}}(R)-\varepsilon)}\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] && (383)\\
&\quad\le \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right] + 2^{n(E_{\mathrm{trc}}(R)-\varepsilon)}2^{-n(E_{\mathrm{rce}}(R)-\varepsilon/2)} && (384)\\
&\quad= \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right] + 2^{-n\varepsilon/2} && (385)
\end{aligned}$$

for n sufficiently large, where (383) follows from Markov's inequality, and (384) follows from $\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] = 2^{-nE_{\mathrm{rce}}(R)}$, so $\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] \le 2^{-n(E_{\mathrm{rce}}(R)-\varepsilon/2)}$ for n sufficiently large. Now, observe that

$$\begin{aligned}
&\mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right] \\
&\quad= \mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)} - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right] && (386)\\
&\quad= \mathbb{P}\left[-\big(P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\big) > \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] - 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right]. && (387)
\end{aligned}$$

Now, since $\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] = 2^{-nE_{\mathrm{rce}}(R)} = 2^{-nE_{\mathrm{trc}}(R)}$ by (380), so $\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] - 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)} > 0$ for n sufficiently large. It follows from (387) that

$$\begin{aligned}
&\mathbb{P}\left[P_{\mathrm{e}}(\mathcal{C}_n) < 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right] \\
&\quad\le \mathbb{P}\left[\left|P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right| > \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] - 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right] && (388)\\
&\quad\le \frac{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}{\left(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] - 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)}\right)^2} && (389)\\
&\quad\doteq \frac{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}{\left(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right)^2} && (390)\\
&\quad\to 0, && (391)
\end{aligned}$$

where (390) follows Markov's inequality and the fact that $\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] - 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)} \doteq 2^{-nE_{\mathrm{rce}}(R)} - 2^{-n(E_{\mathrm{trc}}(R)+\varepsilon)} = \Theta\big(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\big)$ [29, Eq. (28)], and (391) follows from (379). From (385) and (391), we obtain the following result, which is equivalent to (25):

$$\mathbb{P}\left[\left| -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) - E_{\mathrm{trc}}(R)\right| > \varepsilon\right] \to 0, \tag{392}$$

J. Proof of Theorem 10

Under the condition $E_{\mathrm{trc}}(R) > E_{\mathrm{rce}}(R)$, it holds by Theorem 9

$$\frac{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}\, P_{\mathrm{e}}(\mathcal{C}_n)}} \to 0. \tag{393}$$

Now, assume that

$$\frac{P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}} \xrightarrow{\mathrm{(d)}} \mathcal{N}(0,1). \tag{394}$$

Then, from (393) and (394) and Slutsky's theorem [24, p. 334], it holds that

$$\frac{P_{\mathrm{e}}(\mathcal{C}_n)}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}} \xrightarrow{\mathrm{(d)}} \mathcal{N}(0,1), \tag{395}$$

which is a contradiction since the LHS of (395) is a non-negative random variable.

K. Proof of Corollary 1

First, if $\liminf_{n\to\infty} \frac{\mathbb{E}[P_{\mathrm{e}}^2(\mathcal{C}_n)]}{\left(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right)^2} > 1$, then it holds that

$$\nu = \limsup_{n\to\infty} \frac{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}} < \infty. \tag{396}$$

Then, for n sufficiently large, we have

$$\begin{aligned}
\frac{P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}} &\ge \frac{P_{\mathrm{e}}(\mathcal{C}_n)}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}} - \nu && (397)\\
&\ge -\nu, && (398)
\end{aligned}$$

which implies

$$\frac{P_{\mathrm{e}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}(\mathcal{C}_n))}} \overset{\mathrm{(d)}}{\not\to} \mathcal{N}(0,1). \tag{399}$$

Hence, by contradiction, the condition (26) implies that

$$\frac{\mathbb{E}[P_{\mathrm{e}}^2(\mathcal{C}_n)]}{\left(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\right)^2} \to 1. \tag{400}$$

Thanks to (400) we can apply Theorem (8), from which the statement of Corollary 1 follows.

L. Proof of Theorem 11

Let $Y_{ij} = \mathbb{P}\big[\{\boldsymbol{X}_i \to \boldsymbol{X}_j\}\big] - \mathbb{E}\big[\mathbb{P}\big[\{\boldsymbol{X}_i \to \boldsymbol{X}_j\}\big]\big]$ for $i, j \in [M_n] \times [M_n]$. Then, we can write

$$P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)] = \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j \neq i} Y_{ij}$$
$$= \frac{2}{M_n} \sum_{i=1}^{M_n} \sum_{i<j\leq M_n} Y_{ij}. \tag{401}$$

For i.i.d. random coding ensembles, $\{Y_{ij}\}_{1\leq i<j\leq M_n}$ are pairwise independent and identically distributed by the symmetry of the random codebook ensemble. Hence, we have, after some algebra, that

$$\mathrm{Var}\left(\sum_{i=1}^{M_n} \sum_{j \neq i} Y_{ij}\right) = 2M_n(M_n-1)\gamma^2. \tag{402}$$

Hence, by [25, Th. 3.6] with $D \leq 2(M_n-1)$, we have

$$d_W\left(\frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n) - \mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)]}{\sqrt{\mathrm{Var}(P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}}, Z\right)$$
$$\leq \frac{4(M_n-1)^2}{(2M_n(M_n-1)\gamma^2)^{3/2}}\left(\frac{M_n(M_n-1)}{2}\right)\mathbb{E}\big[|Y_{12}|^3\big]$$
$$+ \frac{\sqrt{28}(2(M_n-1))^{3/2}}{\sqrt{\pi}(2M_n(M_n-1)\gamma^2)}\sqrt{\frac{M_n(M_n-1)}{2}\mathbb{E}[|Y_{12}|^4]} \tag{403}$$
$$\leq \frac{M_n}{\gamma^3}\mathbb{E}[|Y_{12}|^3] + \sqrt{\frac{28}{\pi}}\sqrt{M_n \mathbb{E}[|Y_{12}|^4]}, \tag{404}$$

which tends to zero if conditions (30) happen simultaneously.

M. Proof of Theorem 12

We first state two auxiliary lemmas.

Lemma 18: If $Z \sim \mathcal{N}(0,1)$, then for any random variable T, it holds that

$$d_K(T, Z) \leq 2(8\pi)^{-1/4}\sqrt{\tilde{d}_{W,\mathrm{mod}}(T, Z)}. \tag{405}$$

Proof: The proof is similar to the first part of the proof of Lemma 14 in Appendix B.2, so we omit this proof. ■

Lemma 19: If T is a r.v. such that $\mathbb{E}[T] = 0$ and $\mathrm{Var}(T) = 1$, and Z has the standard normal distribution, then

$$d_K(T, Z) < 14(8\pi)^{-1/4}\sqrt{\mathbb{E}[|T|] + \mathbb{E}\big[|T^2 - 1|\big]}. \tag{406}$$

Proof: Appendix C.1. ■

Theorem 12 is a direct application of Lemma 19 by setting $T = g_n(P_{\mathrm{e}}(\mathcal{C}_n))$, gives a criterion for the convergence in distribution of the error exponent and any function of the error probability, in general.

VI. Conclusions

In this paper, we have derived the typical error exponent for the i.i.d. and constant composition random codebook ensembles. We have shown that the random error exponent converges in probability to the typical error exponent of the DMC channel. While this convergence seems plausible for broader families of channels, like finite-state channels, formally proving the result remains an open question. By modifying the Wasserstein metric in Stein's method, we have also shown that the normalized error exponents converge in distribution to a standard Gaussian for a sub-exponential number of codewords or a Gaussian-like distribution at for a constant number of codewords. An related open question is to investigate the convergence in distribution of the normalized error exponent at positive rates.

Appendix A

A.1. Proof of Lemma 1

Since $U_n \xrightarrow{(\mathrm{d})} U$, by the Skorokhod's representation theorem [24, Th. 25.6], there exists a probability space $(\Omega, \mathcal{F}, P)$, a sequence of random variables $\{V_n\}_{n\in\mathbb{N}}$ and a random variable V such that $V_n \sim U_n \ \forall n$ and $V \sim U$ such that $V_n \xrightarrow{(\mathrm{a.s.})} V$ on $(\Omega, \mathcal{F}, P)$. Now, for the given $\varepsilon \in (0,1)$ and any $\delta \in (0, \varepsilon)$, we have

$$\mathbb{E}_P[|V_n^{2+\delta}|] \leq \big(\mathbb{E}_P[|V_n|^{2+\varepsilon}]\big)^{(2+\delta)/(2+\varepsilon)} \tag{407}$$
$$= \big(\mathbb{E}_P[|U_n|^{2+\varepsilon}]\big)^{(2+\delta)/(2+\varepsilon)} \tag{408}$$
$$< L^{(2+\delta)/(2+\varepsilon)} < \infty, \tag{409}$$

where (407) follows from the concavity of the function $f(x) = x^{(2+\delta)/(2+\varepsilon)}$ for any $\varepsilon \in (0,1)$, (408) follows from $U_n \sim V_n$ while (409) follows from the hypothesis of the lemma. From (409), it follows that V_n and V_n^2 are uniformly integrable on $(\Omega, \mathcal{F}, P)$ [24, p. 216 and p. 218 (16.128)]. Hence, we have that

$$\mathbb{E}[U] = \mathbb{E}_P[V] \tag{410}$$
$$= \lim_{n\to\infty} \mathbb{E}_P[V_n] \tag{411}$$
$$= \lim_{n\to\infty} \mathbb{E}[U_n], \tag{412}$$

where (411) follows from [24, Theorem 16.14], and (412) follows from $U_n \sim V_n$.

Similarly, we also have

$$\mathbb{E}[U^2] = \mathbb{E}_P[V^2] \tag{413}$$
$$= \lim_{n\to\infty} \mathbb{E}_P[V_n^2] \tag{414}$$
$$= \lim_{n\to\infty} \mathbb{E}[U_n^2]. \tag{415}$$

Hence, we obtain

$$\mathrm{Var}(V_n) = \mathbb{E}[V_n^2] - \big(\mathbb{E}[V_n]\big)^2 \tag{416}$$
$$\to \mathrm{Var}_P(V) \tag{417}$$

as $n \to \infty$.

From (411) and (417), we obtain

$$\frac{V_n - \mathbb{E}_P[V_n]}{\sqrt{\mathrm{Var}_P(V_n)}} \xrightarrow{(\mathrm{a.s.})} \frac{V - \mathbb{E}_P[V]}{\sqrt{\mathrm{Var}_P(V)}}. \tag{418}$$

Since $U_n \sim V_n$ and $U \sim V$, from (418), we obtain

$$\frac{U_n - \mathbb{E}[U_n]}{\sqrt{\mathrm{Var}(U_n)}} \xrightarrow{(\mathrm{d})} \frac{U - \mathbb{E}[U]}{\sqrt{\mathrm{Var}(U)}}. \tag{419}$$

A.2. Proof of Lemma 3

First, we prove that for any $\alpha > 1$ and $\lambda > 0$, the following holds:

$$\mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{\alpha n}}\big]^{\frac{\alpha n}{\lambda}} \le \mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big]^{\frac{n}{\lambda}}. \tag{420}$$

Indeed, let $r = \frac{\lambda}{\alpha n}$, $p = \frac{\lambda}{n}$, and $q = \frac{\lambda}{(\alpha-1)n}$, satisfying $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ and $p, q, r \in (0, \infty)$ if $\alpha > 1$. By applying the generalized Hölder's inequality [28, p. 140], we have

$$\big(\mathbb{E}[P_{\rm e}(\mathcal{C}_n)^r]\big)^{\frac{1}{r}} \le \big(\mathbb{E}[P_{\rm e}(\mathcal{C}_n)^p]\big)^{\frac{1}{p}} \big(\mathbb{E}[1^q]\big)^{\frac{1}{q}} \tag{421}$$

$$= \big(\mathbb{E}[P_{\rm e}(\mathcal{C}_n)^p]\big)^{\frac{1}{p}}, \tag{422}$$

implying that (420) holds. Since (420) holds for any $\alpha > 1$, we have

$$\mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big]^{\frac{n}{\lambda}} \ge \limsup_{\alpha\to\infty} \mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{\alpha n}}\big]^{\frac{\alpha n}{\lambda}} \tag{423}$$

$$= 2^{\mathbb{E}[\log P_{\rm e}(\mathcal{C}_n)]} \tag{424}$$

where (424) follows from the identity $\mathbb{E}[\log X] = \lim_{x\to\infty} \log \mathbb{E}[X^{\frac{1}{x}}]^x$ for any RV $X > 0$ which is not a function of x. From the definition of $E_{\rm trc}(R,Q)$ in (9) and the definition of limit, we have that for every $\epsilon > 0$ there exists an $n_0(\varepsilon)$ such that for $n > n_0(\varepsilon)$,

$$\Big| -\frac{1}{n}\mathbb{E}[\log P_{\rm e}(\mathcal{C}_n)] - E_{\rm trc}(R)\Big| < \epsilon. \tag{425}$$

Therefore, from (424) we have that

$$\mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big]^{\frac{n}{\lambda}} \ge 2^{-n(1-\varepsilon)E_{\rm trc}(R,Q)}, \quad \forall n \ge n_0(\varepsilon). \tag{426}$$

Thus, from (426) and (420), it holds that

$$\liminf_{n\to\infty} \mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big] \ge 2^{-\lambda(1-\varepsilon)E_{\rm trc}(R,Q)} \tag{427}$$

for all $\varepsilon > 0$. This means that

$$\liminf_{n\to\infty} \mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big] \ge 2^{-\lambda E_{\rm trc}(R,Q)} \tag{428}$$

by letting $\varepsilon \to 0$.

Now, by the concavity of the function $g(x) := x^{\frac{\lambda}{n}}$ on $(0, \infty)$, we have by Jensen's inequality that

$$\limsup_{n\to\infty} \mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big] \le \limsup_{n\to\infty} \mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)\big]^{\frac{\lambda}{n}} \tag{429}$$

$$\le 2^{-\lambda E_{\rm rce}(R,Q)}. \tag{430}$$

where (430) follows from the fact that $\mathbb{E}[P_{\rm e}(\mathcal{C}_n)] \overset{.}{\le} 2^{-nE_{\rm rce}(R,Q)}$ [3, Theorem 1], [30, Theorem 8.7].

Finally, from (428) and (430), under the condition that $E_{\rm trc}(R,Q) = E_{\rm rce}(R,Q)$, it holds that

$$\lim_{n\to\infty} \mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)^{\frac{\lambda}{n}}\big] = 2^{-\lambda E_{\rm rce}(R,Q)}. \tag{431}$$

A.3. Proof of Lemma 4

The upper bound follows from Bhattacharyya bound. Now, by [29, Eq. (28)], it holds that

$$\mathbb{E}\big[P_{\rm e}(\mathcal{C}_n)\big] \doteq 2^{-nE_{\rm rce}(R,Q)} \tag{432}$$

for $R < R_{\rm crit}$. In addition, at this range of rate, the Bhattacharyya bound achieves the Gallager's random coding bound $E_{\rm rce}(R,Q)$. Hence, from (432), we have

$$\mathbb{E}\big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big] \doteq 2^{-nE_{\rm rce}(R,Q)} \tag{433}$$

for $R < R_{\rm crit}$, where $P_{\rm e}^{\rm ub}(\mathcal{C}_n)$ is the union bound on $P_{\rm e}(\mathcal{C}_n)$.

Now, for all rate $R < R_{\rm crit}$, $E_{\rm rce}(R,Q) = R_0(Q) - R$, where R_0 is the cut-off rate corresponding to the underlying distribution Q, i.e.,

$$R_0(Q) = -\log\bigg(\sum_y \bigg(\sum_x Q(x)\sqrt{W(y|x)}\bigg)^2\bigg). \tag{434}$$

Let $Q_X = Q'_X = Q$. By using standard KKT conditions for convex optimization, it is not hard to prove that

$$R_0(Q) = \min_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X})} D(P_{XX'}\|Q_XQ'_X) + \sum_{x,x'} P_{XX'}(x,x')d_{\rm B}(x,x'). \tag{435}$$

Now, recall the definition of K_2 in (153). From (435), we obtain

$$\mathbb{E}\big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big] = M_n \sum_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X})} 2^{-nD(P_{XX'}\|Q_{XX'})} \times 2^{-n\sum_{x,x'}P_{XX'}(x,x')d_{\rm B}(x,x')}. \tag{436}$$

Now, let $\mathcal{N}(P_{XX'})$ be the number of codeword pairs which have the same join type $P_{XX'}$. Then, it holds that

$$\mathcal{N}(P_{XX'}) = \sum_{i=1}^{M_n}\sum_{j\ne i} \mathbb{1}\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\}, \tag{437}$$

which leads to

$$\mathbb{E}[\mathcal{N}(P_{XX'})] = M_n(M_n-1)2^{-nD(P_{XX'}\|Q_{XX'})}. \tag{438}$$

From (436) and (438), we have

$$(M_n-1)\mathbb{E}\big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big] = \sum_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X})} \mathbb{E}[\mathcal{N}(P_{XX'})]2^{-n\sum_{x,x'}P_{XX'}(x,x')d_{\rm B}(x,x')}. \tag{439}$$

On the other hand, observe that

$$(M_n-1)\mathbb{E}\big[P_{\rm e}^{\rm ub}(\mathcal{C}_n)\big] = \mathbb{E}\bigg[\sum_{i=1}^{M_n}\sum_{j\ne i}\mathbb{P}\big(\boldsymbol{X}_i \to \boldsymbol{X}_j\big)\bigg]. \tag{440}$$

From (439) and (440), we obtain

$$\mathbb{E}\bigg[\sum_{i=1}^{M_n}\sum_{j\ne i}\mathbb{P}\big(\boldsymbol{X}_i \to \boldsymbol{X}_j\big)\bigg] = \sum_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X})} 2^{-nD(P_{XX'}\|Q_{XX'})} \times 2^{-n\sum_{x,x'}P_{XX'}(x,x')d_{\rm B}(x,x')}. \tag{441}$$

Since (441) holds for all random i.i.d. codebook ensembles, hence for any $P_{XX'} \in \mathcal{P}(\mathcal{X}\times\mathcal{X})$, by choosing a sub-random codebook ensemble which contains all the codewords with the same joint type $P_{XX'}$, we obtain

$$\mathbb{P}\Big[\boldsymbol{X}_i \to \boldsymbol{X}_j \Big| (\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\Big] = 2^{-n\sum_{x,x'} P_{XX'}(x,x') d_{\mathrm{B}}(x,x')}. \tag{442}$$

A.4. Proof of Lemma 5

Observe that

$$\begin{aligned}
\mathbb{P}[\mathcal{V}_n^c] &= \mathbb{P}\Bigg[\sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') > 2R} \mathcal{N}(P_{XX'}) \geq 1\Bigg] && (443)\\
&\leq \mathbb{E}\Bigg[\sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') > 2R} \mathcal{N}(P_{XX'})\Bigg] && (444)\\
&= \sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') > 2R} \sum_{i=1}^{M_n} \sum_{j\neq i} \mathbb{P}\big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\big] && (445)\\
&\leq \sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') > 2R} \sum_{i=1}^{M_n} \sum_{j\neq i} 2^{-nD(P_{XX'}\|Q_X Q_X')} && (446)\\
&\dot{\leq} 2^{2nR} \sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') > 2R} 2^{-nD(P_{XX'}\|Q_X Q_X')} && (447)\\
&\dot{\leq} 2^{2nR} \sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') > 2R} 2^{-n(2R+\alpha(R))} && (448)\\
&\dot{\leq} 2^{-n\alpha(R)} && (449)
\end{aligned}$$

for some $\alpha(R) > 0$, where (447) follows from $M_n \doteq 2^{nR}$, and (449) follows from the fact that the number of possible n-joint types on $\mathcal{X} \times \mathcal{X}$ is sub-exponential in n.

A.5. Proof of Lemma 6

Define

$$\tilde{V}_{ij} = \sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') \leq 2R-\nu} \mathbb{1}\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\} \times g_n(P_{XX'}). \tag{450}$$

Then, we have

$$\begin{aligned}
D_n &= \frac{1}{M_n} \sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') \leq 2R-\nu} \mathcal{N}(P_{XX'}) \times g_n(P_{XX'}) && (451)\\
&= \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j\neq i} \sum_{P_{XX'}: D(P_{XX'}\|Q_X Q_X') \leq 2R-\nu} \mathbb{1}\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\} g_n(P_{XX'}) && (452)\\
&= \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j\neq i} \tilde{V}_{ij}, && (453)
\end{aligned}$$

where $\{\tilde{V}_{ij}\}_{i,j=1}^{M_n}$ is a sequence of independent random variables. Hence, from (453), we have

$$\mathrm{Var}(D_n) = \frac{1}{M_n^2} \sum_{i=1}^{M_n} \sum_{j\neq i} \mathrm{Var}(\tilde{V}_{ij}). \tag{454}$$

Now, let

$$\mathcal{A}_\nu := \big\{P_{XX'} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}) : D(P_{XX'}\|Q_X Q_X') \leq 2R - \nu\big\} \tag{455}$$

and recall the definition of $F(P_{XX'})$ in (161).

Observe that

$$\begin{aligned}
&\mathrm{Var}(\tilde{V}_{ij})\\
&= \sum_{P_{XX'}\in\mathcal{A}_\nu} \Bigg(\mathbb{E}\Bigg[\Big(\mathbb{1}\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\}\Big)^2\Bigg] - \Bigg(\mathbb{E}\Big[\mathbb{1}\{(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\}\Big]\Bigg)^2\Bigg) g_n^2(P_{XX'}) && (456)\\
&= \sum_{P_{XX'}\in\mathcal{A}_\nu} \mathbb{P}\Big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\Big] \times \Bigg(1 - \mathbb{P}\Big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\Big]\Bigg) g_n^2(P_{XX'}) && (457)\\
&\leq 2^{n\max_{P_{XX'} P_{XX'}\in\mathcal{A}_\nu} -F(P_{XX'})}) \times \sum_{P_{XX'} P_{XX'}\in\mathcal{A}_\nu} g_n(P_{XX'}) \mathbb{P}\Big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\Big], && (458)
\end{aligned}$$

where (458) follows from Lemma 4.

Hence, from (454) and (458), we obtain

$$\begin{aligned}
&\mathrm{Var}(D_n)\\
&= 2^{-2nR} \times 2^{n\max_{P_{XX'}\in\mathcal{A}_\nu} -F(P_{XX'})} \times \sum_{P_{XX'}\in\mathcal{A}_\nu} \sum_{i=1}^{M_n} \sum_{j\neq i} \mathbb{P}\Big[(\boldsymbol{X}_i, \boldsymbol{X}_j) \in \mathcal{T}_n(P_{XX'})\Big] g_n(P_{XX'}) && (459)\\
&= 2^{-2nR} \times 2^{n\max_{P_{XX'}\in\mathcal{A}_\nu} -F(P_{XX'})} \times \Bigg(\sum_{P_{XX'}\in\mathcal{A}_\nu} \mathbb{E}[\mathcal{N}(P_{XX'})] g_n(P_{XX'})\Bigg) && (460)\\
&\doteq 2^{-nR} \times 2^{-n\min_{P_{XX'}\in\mathcal{A}_\nu} F(P_{XX'})} \mathbb{E}[D_n], && (461)
\end{aligned}$$

where (461) follows from the fact that the optimizer of the linear objective function over the convex constraint set is in the boundary of the convex constraint set[5] On the other hand,

[5]We can easily to check this fact by using the Karush–Kuhn–Tucker conditions [31, Sec. 5.5.3]. Note that $d_{\mathrm{B}}(x,x') > 0$ for all x, x'.

from (453), we have

$$\mathbb{E}[D_n]$$
$$= \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\sum_{P_{XX'}\in\mathcal{A}_\nu}\mathbb{P}\big[(\boldsymbol{X}_i,\boldsymbol{X}_j)\in\mathcal{T}_n(P_{XX'})\big]g_n(P_{XX'}) \tag{462}$$
$$= \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\sum_{P_{XX'}\in\mathcal{A}_\nu}2^{-nD(P_{XX'}\|Q_XQ'_X)}g_n(P_{XX'}) \tag{463}$$
$$= (M_n-1)2^{-n\min_{P_{XX'}\in\mathcal{A}_\nu}\left(D(P_{XX'}\|Q_XQ'_X)+F(P_{XX'})\right)}, \tag{464}$$

where (463) follows from [32, p. 2506], and (464) follows from the definition of pairwise error probability given in Lemma 4. Hence, we obtain (51). Using a similar reasoning and the fact that $\mathcal{P}_n(\mathcal{X}\times\mathcal{X})$ is dense in $\mathcal{P}(\mathcal{X}\times\mathcal{X})$, we obtain that

$$\mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)]$$
$$= \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_i\to\boldsymbol{X}_j]\big] \tag{465}$$
$$= 2^{-n\min_{P_{XX'}}\left(D(P_{XX'}\|Q_XQ'_X)+F(P_{XX'})-R\right)}, \tag{466}$$

implying that

$$E_{\mathrm{rce}}(R,Q)+R = \min_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X})} D(P_{XX'}\|Q_XQ'_X)+F(P_{XX'}). \tag{467}$$

Using the Karush–Kuhn–Tucker conditions [31, Sec. 5.5.3], we can show equation (468) at the top of next page, where $P^*_{XX'}$ minimizes $\min_{P_{XX'}\in\mathcal{P}} D(P_{XX'}\|Q_XQ_{X'})+F(P_{XX'})-R$. From (467) and (468), it holds that

$$2R-\nu+F(P^*_{XX'})\geq E_{\mathrm{rce}}(R,Q)+R. \tag{469}$$

From (464) and (468), we obtain (51). Furthermore, from (461) and (51), we obtain

$$\frac{\mathrm{Var}(D_n)}{\big(\mathbb{E}[D_n]\big)^2} \;\dot{\leq}\; \begin{cases} 2^{-n\nu}, & D(P^*_{XX'}|Q_XQ'_X)=2R-\nu \\ 2^{-n\min_{P_{XX'}:D(P_{XX'}\|Q_XQ'_X)=2R-\nu}F(P_{XX'})} \times 2^{n(R-E_{\mathrm{rce}}(R,Q))}, & \text{otherwise} \end{cases} \tag{470}$$
$$\leq 2^{-\nu n} \tag{471}$$

where (471) follows from (469).

This concludes our proof of Lemma 6.

A.6. Proof of Lemma 7

From the definitions of $\mathcal{N}(P_{XX'})$ in Lemma 5 and $P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)$ in (15), we have

$$P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)=\frac{1}{M_n}\sum_{P_{XX'}}\mathcal{N}(P_{XX'})g_n(P_{XX'}). \tag{472}$$

First, we consider the case $R>0$. Take an arbitrary ν such that $0<\nu\leq 2R$. Let

$$B_n=\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{X})}\mathcal{N}(P_{XX'})g_n(P_{XX'}), \tag{473}$$
$$\tilde{D}_n=\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{X}):D(P_{XX'}\|Q_XQ'_X)\leq 2R}\mathcal{N}(P_{XX'}) \times g_n(P_{XX'}), \tag{474}$$
$$D_n=\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{X}):D(P_{XX'}\|Q_XQ'_X)\leq 2R-\nu}\mathcal{N}(P_{XX'}) \times g_n(P_{XX'}). \tag{475}$$

Since $\mathbb{E}[\tilde{D}_n]-\mathbb{E}[D_n]\to 0$ as $\nu\to 0$ and that $\mathbb{E}[\tilde{D}_n]$ and $\mathbb{E}[D_n]$ are exponentially decaying in n, for ν small enough, it holds that

$$\mathbb{E}[D_n]\leq\mathbb{E}[\tilde{D}_n]\leq\mathbb{E}[D_n]2^{\varepsilon n/2}. \tag{476}$$

Recall the typical set $\mathcal{V}_n$ defined in Lemma 5. For any given $\varepsilon>0$, observe that

$$\mathbb{P}\Bigg[\bigg|-\frac{\log P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)}{n} + \frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_\nu}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_n(P_{XX'})\bigg)\bigg|>\varepsilon\Bigg] \tag{477}$$
$$=\mathbb{P}\bigg[\bigg|\log\frac{B_n}{\mathbb{E}[\tilde{D}_n]}\bigg|>\varepsilon n\bigg] \tag{478}$$
$$\leq\mathbb{P}\bigg[\bigg|\log\frac{B_n}{\mathbb{E}[\tilde{D}_n]}\bigg|>\varepsilon n\bigg|\mathcal{V}_n\bigg]\mathbb{P}[\mathcal{V}_n]+\mathbb{P}[\mathcal{V}_n^c] \tag{479}$$
$$=\mathbb{P}\bigg[\bigg|\log\frac{\tilde{D}_n}{\mathbb{E}[\tilde{D}_n]}\bigg|>\varepsilon n\bigg|\mathcal{V}_n\bigg]\mathbb{P}[\mathcal{V}_n]+\mathbb{P}[\mathcal{V}_n^c] \tag{480}$$
$$\leq\mathbb{P}\bigg[\bigg|\log\frac{\tilde{D}_n}{\mathbb{E}[\tilde{D}_n]}\bigg|>\varepsilon n\bigg]+\mathbb{P}[\mathcal{V}_n^c] \tag{481}$$
$$=\mathbb{P}\big[\tilde{D}_n>\mathbb{E}[\tilde{D}_n]2^{\varepsilon n}\big]+\mathbb{P}\big[\tilde{D}_n<\mathbb{E}[\tilde{D}_n]2^{-\varepsilon n}\big]+\mathbb{P}[\mathcal{V}_n^c] \tag{482}$$
$$\leq 2^{-\varepsilon n}+\mathbb{P}\big[D_n<\mathbb{E}[\tilde{D}_n]2^{-\varepsilon n}\big]+2^{-n\alpha(R)} \tag{483}$$
$$\leq 2^{-\varepsilon n}+\mathbb{P}\big[D_n<\mathbb{E}[D_n]2^{-(\varepsilon/2)n}\big]+2^{-n\alpha(R)} \tag{484}$$

where (479) follows from $\mathbb{P}(A)\leq\mathbb{P}(A|B)\mathbb{P}(B)+\mathbb{P}(B^c)$, (480) follows from the fact that given $\mathcal{V}_n$, it holds that $B_n=D_n$, (481) follows from $\mathbb{P}(A|B)\mathbb{P}(B)\leq\mathbb{P}(A)$, (483) follows from Markov's inequality and Lemma 5, and (484) follows from $D_n\leq\tilde{D}_n$ and $\mathbb{E}[\tilde{D}_n]\leq\mathbb{E}[D_n]2^{\varepsilon n/2}$ for ν sufficiently small by (476).

Now, we have

$$\mathbb{P}\big[D_n<\mathbb{E}[D_n]2^{-(\varepsilon/2)n}\big]$$
$$\leq\mathbb{P}\big[\big|D_n-\mathbb{E}[D_n]\big|>\mathbb{E}[D_n]\big(1-2^{-(\varepsilon/2)n}\big)\big] \tag{485}$$
$$\dot{\leq}\frac{\mathrm{Var}(D_n)}{\big(\mathbb{E}[D_n]\big)^2} \tag{486}$$
$$\dot{\leq} 2^{-n\nu}, \tag{487}$$

$$\min_{P_{XX'}\in\mathcal{A}_\nu} D(P_{XX'}\|Q_XQ'_X)+F(P_{XX'})=\begin{cases}2R-\nu+F(P^*_{XX'}), & \text{if} \quad D(P^*_{XX'}\|Q_XQ'_X)=2R-\nu,\\ R+E_{\rm rce}(R,Q), & \text{otherwise.}\end{cases} \tag{468}$$

where (487) follows from Lemma 6. From (484) and (487), for any $\varepsilon>0$ and $R>0$, we have

$$\mathbb{P}\bigg[\bigg|-\frac{\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n} + \frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_0}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_n(P_{XX'})\bigg)\bigg|>\varepsilon\bigg] \tag{488}$$

$$\dot{\leq} 2^{-\varepsilon n}+2^{-n\nu}+2^{-n\alpha(R)} \tag{489}$$

where $\mathcal{A}_0$ is defined in (455) at $\nu=0$.

It follows from (489) that

$$\sum_{n=1}^{\infty}\mathbb{P}\bigg[\bigg|-\frac{\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n} + \frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_0}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_n(P_{XX'})\bigg)\bigg|>\varepsilon\bigg] < \infty. \tag{490}$$

Hence, by Borel-Cantelli's lemma [24, Theorem 4.3], we have

$$\frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_0}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_N(P_{XX'})\bigg) - \frac{\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n}\xrightarrow{\text{(a.s.)}}0 \tag{491}$$

On the other hand, we have

$$\bigg|-\frac{\log P_{\rm e}^{\rm ub}(\mathcal{C}_n))}{n} + \frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_0}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_n(P_{XX'})\bigg)\bigg| \leq -\frac{\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n} + \bigg|-\frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_0}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_n(P_{XX'})\bigg)\bigg| \tag{492}$$

$$\leq E_{\rm sp}(R)+\bigg|-\frac{1}{n}\log\bigg(\frac{\mathbb{E}[D_n]}{M_n}\bigg)\bigg|, \tag{493}$$

$$\leq E_{\rm sp}(R)+R+\min_{P_{XX'}:D(P_{XX'}\|Q_XQ'_X)=2R}F(P_{XX'}) \tag{494}$$

$$\leq E_{\rm sp}(R)+R+D_{\rm b}<\infty, \tag{495}$$

where (493) follows from [30, Theorem 8.11], (494) follows from Lemma 6, where (495) follows with the fact that $d_{\rm B}(x,x')\leq D_{\rm b}<\infty$ for all x,x' by the condition (46).

From (491), (495), and the bounded convergence theorem [24, Theorem 5.4], we have

$$\lim_{n\to\infty}\mathbb{E}\bigg[-\frac{\log P_{\rm e}^{\rm ub}(\mathcal{C}_n)}{n} + \frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_0}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_n(P_{XX'})\bigg)\bigg]=0. \tag{496}$$

Now, by Lemma 6, we have

$$\lim_{n\to\infty}-\frac{1}{n}\log\bigg(\frac{1}{M_n}\sum_{P_{XX'}\in\mathcal{A}_0}\mathbb{E}\big[\mathcal{N}(P_{XX'})\big]g_n(P_{XX'})\bigg) = \min_{P_{XX'}:D(P_{XX'}\|Q_XQ'_X)\leq 2R}\big(D(P_{XX'}\|Q_XQ'_X) + F(P_{XX'})-R\big). \tag{497}$$

Hence, we obtain (54) from (496) and (497). Note that (55) can be achieved from (54) by using (468) with $\nu=0$.

A.7. Proof of Lemma 8

For any $\varepsilon>0$, by Chebyshev's inequality, we have

$$\mathbb{P}\bigg[\Big|P_{\rm e}^{\rm ub}(\mathcal{C}_n)-\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]\Big|\geq 2^{-n\varepsilon}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]\bigg] \leq \frac{{\rm Var}(P_{\rm e}^{\rm ub}(\mathcal{C}_n))}{2^{-2n\varepsilon}\big(\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]\big)^2}. \tag{498}$$

On the other hand, we have

$${\rm Var}(P_{\rm e}^{\rm ub}(\mathcal{C}_n)) \dot{\leq} 2^{-nR}2^{-n\min_{P_{XX'}:D(P_{XX'}\|Q_XQ_{X'})=2R}F(P_{XX'})}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)] \tag{499}$$

$$=2^{-nE_{\rm trc}^{\rm ub}(R,Q)}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]. \tag{500}$$

where (499) follows from (461), (500) follows from (49) with $\nu=0$ and $E_{\rm trc}^{\rm ub}(R,Q)>E_{\rm rce}(R,Q)$ so $E_{\rm trc}^{\rm ub}(R,Q)=\min_{P_{XX'}\in\mathcal{P}(\mathcal{X}\times\mathcal{X}):D(P_{XX'}\|Q_XQ_{X'})=2R}F(P_{XX'})$.

On the other hand, for $R<R_{\rm crit}$, we have

$$\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]\doteq 2^{-nE_{\rm rce}(R,Q)}. \tag{501}$$

From (498), (500), and (501), we obtain

$$\mathbb{P}\bigg[\Big|P_{\rm e}^{\rm ub}(\mathcal{C}_n)-\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]\Big|\geq 2^{-n\varepsilon}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]\bigg]. \tag{502}$$

Now, for the case $E_{\rm trc}^{\rm ub}(R,Q)>E_{\rm rce}(R,Q)$, observe that

$$\mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n)\geq\frac{1}{2}2^{-n[E_{\rm trc}^{\rm ub}(R,Q)-\epsilon]}\right] + \mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n)\leq 2^{-n[E_{\rm trc}^{\rm ub}(R,Q)+\epsilon]}\right] \tag{503}$$

$$\leq\frac{1}{n^{1+\kappa}}+\mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n)\leq 2^{-n[E_{\rm trc}^{\rm ub}(R,Q)+\epsilon]}\right] \tag{504}$$

where (504) follows from [15, Eq. (22)] with $\gamma_n = n^{1+\kappa'}$ for some $\kappa' > 0$. Next, we bound the second term in (504) for large values of n.

Define

$$\mathcal{A} := \left\{ \left| P_{\rm e}^{\rm ub}(\mathcal{C}_n) - \mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)] \right| \geq 2^{-n\varepsilon}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)] \right\}. \tag{505}$$

Then, on $\mathcal{A}^c$, we have

$$P_{\rm e}^{\rm ub}(\mathcal{C}_n) \geq \mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)] - 2^{-n\varepsilon}\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)] \tag{506}$$
$$\doteq 2^{-nE_{\rm rce}(R,Q)}. \tag{507}$$

Hence, we have

$$\mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n) \leq 2^{-n[E_{\rm trc}^{\rm ub}(R,Q)+\epsilon]}\right] \tag{508}$$
$$\leq \mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n) \leq 2^{-n[E_{\rm trc}^{\rm ub}(R,Q)+\epsilon]} \big| \mathcal{A}^c\right] + \mathbb{P}(\mathcal{A}) \tag{509}$$
$$\leq \mathbf{1}\left\{2^{-nE_{\rm rce}(R,Q)} \leq 2^{-n[E_{\rm trc}^{\rm ub}(R,Q)+\epsilon]}\right\} + 2^{-n(E_{\rm trc}(R,Q)-E_{\rm rce}(R,Q)-2\varepsilon)} \tag{510}$$
$$= 2^{-n(E_{\rm trc}^{\rm ub}(R,Q)-E_{\rm rce}(R,Q)-2\varepsilon)}, \tag{511}$$

where (509) follows from $\mathbb{P}(A) = \mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|E^c)\mathbb{P}(E^c) \leq \mathbb{P}(A|E) + \mathbb{P}(E^c)$ for any set E, (510) follows from (502) and (507), and (511) follows from $E_{\rm trc}^{\rm ub}(R,Q) > E_{\rm rce}(R,Q)$.

By choosing $\varepsilon := (E_{\rm trc}(R,Q) - E_{\rm rce}(R,Q))/4$, from (504) and (511), we obtain

$$\mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n) \geq \frac{1}{2}2^{-n[E_{\rm trc}^{\rm ub}(R,Q)-\epsilon]}\right] + \mathbb{P}\left[P_{\rm e}^{\rm ub}(\mathcal{C}_n) \leq 2^{-n[E_{\rm trc}^{\rm ub}(R,Q)+\epsilon]}\right]$$
$$\leq \frac{1}{n^{1+\kappa}} + 0 + 2^{-n(E_{\rm trc}(R,Q)-E_{\rm rce}(R,Q))/2} \tag{512}$$
$$\doteq \frac{1}{n^{1+\kappa}}. \tag{513}$$

From (513), our proof is concluded.

A.8. Proof of Lemma 9

Observe that equations (514)–(520) at the top of next page hold, where (517) follows from Caen's inequality in Lemma 2 by, for each fixed i, setting $\mathcal{I}_i = \{j \in [M] \setminus \{i\} : j \neq i\}, A_j^{(i)} = \{\boldsymbol{X}_i \to \boldsymbol{X}_j\}$ with the probability measure defined as $\mathbb{P}(A_j^{(i)}) = \mathbb{E}[\mathbb{E}[\mathbb{1}\{\boldsymbol{X}_i \to \boldsymbol{X}_j\}]] = \mathbb{E}[\mathbb{P}[\boldsymbol{X}_i \to \boldsymbol{X}_j]]$, where the inner expectation is over the channel randomness and the outer one is over the random codebook ensemble. This is the probability of event $\{\boldsymbol{X}_i \to \boldsymbol{X}_j\}$ on the a product probability space generated from channel statistics and random codebook generations. By the symmetry of the codebook generation, we have that $\mathbb{P}(A_j^{(i)}) = \mathbb{E}[\mathbb{P}[\boldsymbol{X}_i \to \boldsymbol{X}_j]] = \mathbb{E}[\mathbb{P}[\boldsymbol{X}_1 \to \boldsymbol{X}_2]] = \mathbb{P}(A_2^{(1)})$ for all $j \neq i$. From (520), it holds that

$$1 \leq \frac{\mathbb{E}[P_{\rm e}^{\rm ub}(\mathcal{C}_n)]}{\mathbb{E}[P_{\rm e}(\mathcal{C}_n)]} \tag{521}$$
$$\leq 1 + (M_n - 2)\frac{\mathbb{E}[\mathbb{P}[\{\boldsymbol{X}_1 \to \boldsymbol{X}_2\} \cap \{\boldsymbol{X}_1 \to \boldsymbol{X}_3\}]]}{\mathbb{E}[\mathbb{P}[\boldsymbol{X}_1 \to \boldsymbol{X}_2]]}. \tag{522}$$

Recall the definition of $d_{\rm B}(x,x')$ in (45). Assume that $\boldsymbol{x}_1 \in \mathcal{T}_n(P_X)$ for some $P_X \in \mathcal{P}_n(\mathcal{X})$, which is a fixed vector. Then, given $(\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathcal{T}_n(P_{XX'})$ and $(\boldsymbol{x}_1,\boldsymbol{x}_3) \in \mathcal{T}_n(P_{XX"})$ where $P_{XX'} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X})$ and $P_{XX"} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X})$, it holds that

$$\mathbb{P}\{\{\boldsymbol{x}_1 \to \boldsymbol{x}_2\} \cap \{\boldsymbol{x}_1 \to \boldsymbol{x}_3\} | (\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathcal{T}_n(P_{XX'}), (\boldsymbol{x}_1,\boldsymbol{x}_3) \in \mathcal{T}_n(P_{XX"})\}$$
$$\leq \min\left\{\mathbb{P}\{\boldsymbol{x}_1 \to \boldsymbol{x}_2 | (\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathcal{T}_n(P_{XX'})\}, \mathbb{P}\{\boldsymbol{x}_1 \to \boldsymbol{x}_3 | (\boldsymbol{x}_1,\boldsymbol{x}_3) \in \mathcal{T}_n(P_{XX"})\}\right\} \tag{523}$$
$$\leq \min\left\{2^{-nF(P_{XX'})}, 2^{-nF(P_{XX"})}\right\} \tag{524}$$
$$= 2^{-n\max\left\{F(P_{XX'}, F(P_{XX"})\right\}}, \tag{525}$$

which does not depend on $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$, where (524) follows from Lemma 4. In addition, we have

$$\mathbb{P}\Big[(\boldsymbol{x}_1, \boldsymbol{X}_2) \in \mathcal{T}_n(P_{XX'})\Big]$$
$$= \sum_{\boldsymbol{x}_2} \mathbb{P}(\boldsymbol{x}_2)\mathbb{1}\{(\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathcal{T}_n(P_{XX'})\} \tag{526}$$
$$= 2^{-n\left(H(P'_X)+D(P_{X'}\|Q)\right)}\sum_{\boldsymbol{x}_2}\mathbb{1}\{(\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathcal{T}_n(P_{XX'})\} \tag{527}$$
$$= 2^{-n\left(H(P'_X)+D(P_{X'}\|Q)\right)}\frac{|\mathcal{T}_n(P_{XX'}|}{|\mathcal{T}_n(P_X)|} \tag{528}$$
$$= 2^{-n\left(I_P(X;X')+D(P'_X\|Q)\right)}, \tag{529}$$

where (527) and (529) follow from [32, p. 2056]. Similarly, we also have

$$\mathbb{P}\Big[(\boldsymbol{x}_1, \boldsymbol{X}_3) \in \mathcal{T}_n(P_{XX"})\Big] = 2^{-n\left(I_P(X;X")+D(P"_X\|Q)\right)}. \tag{530}$$

Hence, we have

$$\mathbb{P}\Big[\{(\boldsymbol{x}_1,\boldsymbol{X}_2) \in \mathcal{T}_n(P_{XX'})\} \cap \{(\boldsymbol{x}_1,\boldsymbol{X}_3) \in \mathcal{T}_n(P_{XX"})\}\Big| \boldsymbol{X}_1 = \boldsymbol{x}_1\Big]$$
$$= \mathbb{P}\Big[\{(\boldsymbol{x}_1,\boldsymbol{X}_2) \in \mathcal{T}_n(P_{XX'})\} \cap \{(\boldsymbol{x}_1,\boldsymbol{X}_3) \in \mathcal{T}_n(P_{XX"})\}\Big] \tag{531}$$
$$= \mathbb{P}\Big[(\boldsymbol{x}_1,\boldsymbol{X}_2) \in \mathcal{T}_n(P_{XX'})\Big]\mathbb{P}\Big[(\boldsymbol{x}_1,\boldsymbol{X}_3) \in \mathcal{T}_n(P_{XX"})\Big] \tag{532}$$
$$= 2^{-n\left(I_P(X;X')+D(P'_X\|Q)\right)}2^{-n\left(I_P(X;X")+D(P"_X\|Q)\right)} \tag{533}$$
$$= 2^{-n(I_P(X;X')+I_P(X;X")+D(P'_X\|Q)+D(P_{X"}\|Q))}, \tag{534}$$

$$
\begin{aligned}
\mathbb{E}[P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)] &\geq \mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)] && (514)\\
&= \frac{1}{M_n}\sum_{i=1}^{M_n}\mathbb{E}\bigg[\mathbb{P}\bigg(\bigcup_{j\neq i}\{\boldsymbol{X}_i\to\boldsymbol{X}_j\}\bigg)\bigg] && (515)\\
&= \frac{1}{M_n}\sum_{i=1}^{M_n}\mathbb{E}\bigg[\mathbb{E}\bigg[\mathbb{1}\bigg\{\bigcup_{j\neq i}\{\boldsymbol{X}_i\to\boldsymbol{X}_j\}\bigg\}\bigg]\bigg] && (516)\\
&\geq \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\frac{\big(\mathbb{E}\big[\mathbb{E}[\mathbb{1}\{\boldsymbol{X}_i\to\boldsymbol{X}_j\}]\big]\big)^2}{\mathbb{E}[\mathbb{E}[\mathbb{1}\{\boldsymbol{X}_i\to\boldsymbol{X}_j\}]]+\sum_{k\neq i,j}\mathbb{E}[\mathbb{E}[\mathbb{1}\{\{\boldsymbol{X}_i\to\boldsymbol{X}_j\}\cap\{\boldsymbol{X}_i\to\boldsymbol{X}_k\}\}]]} && (517)\\
&= \frac{1}{M_n}\sum_{i=1}^{M_n}\sum_{j\neq i}\frac{\big(\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_i\to\boldsymbol{X}_j]\big]\big)^2}{\mathbb{E}[\mathbb{P}[\boldsymbol{X}_i\to\boldsymbol{X}_j]]+\sum_{k\neq i,j}\mathbb{E}[\mathbb{P}[\{\boldsymbol{X}_i\to\boldsymbol{X}_j\}\cap\{\boldsymbol{X}_i\to\boldsymbol{X}_k\}]} && (518)\\
&= \frac{(M_n-1)\big(\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_1\to\boldsymbol{X}_2]\big]\big)^2}{\mathbb{E}[\mathbb{P}[\boldsymbol{X}_1\to\boldsymbol{X}_2]]+(M_n-2)\mathbb{E}[\mathbb{P}[\{\boldsymbol{X}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{X}_1\to\boldsymbol{X}_3\}]]} && (519)\\
&= \frac{P_{\mathrm{e}}^{\mathrm{ub}}(\mathcal{C}_n)\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_1\to\boldsymbol{X}_2]\big]}{\mathbb{E}\big[\mathbb{P}[\boldsymbol{X}_1\to\boldsymbol{X}_2]\big]+(M_n-2)\mathbb{E}\big[\mathbb{P}[\{\boldsymbol{X}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{X}_1\to\boldsymbol{X}_3\}]\big]}. && (520)
\end{aligned}
$$

---

where (533) follows from (529) and (530). It follows from (525) and (534) that

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}}[\mathbb{P}[\{\boldsymbol{x}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{x}_1\to\boldsymbol{X}_3\}]]\\
&= \sum_{P_{X'|X}}\sum_{P_{X"|X}}\mathbb{E}\bigg[\mathbb{P}\{\{\boldsymbol{x}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{x}_1\to\boldsymbol{X}_3\}\\
&\qquad\bigg|(\boldsymbol{x}_1,\boldsymbol{X}_2)\in\mathcal{T}_{P_{XX'}},(\boldsymbol{x}_1,\boldsymbol{X}_3)\in\mathcal{T}_{P_{XX"}}\}\\
&\qquad\times\mathbb{P}\bigg[\{(\boldsymbol{x}_1,\boldsymbol{X}_2)\in\mathcal{T}_n(P_{XX'})\}\\
&\qquad\cap\{(\boldsymbol{x}_1,\boldsymbol{X}_3)\in\mathcal{T}_n(P_{XX"})\}\bigg|\boldsymbol{X}_1=\boldsymbol{x}_1\bigg]\bigg] && (535)\\
&\leq \sum_{P_{X'|X}}\sum_{P_{X"|X}}2^{-n\max\{F(P_{XX'}),F(P_{XX"})\}}\\
&\qquad\times 2^{-n(I_P(X;X')+I_P(X;X")+D(P'_X\|Q)+D(P"_X\|Q))} && (536)\\
&\leq \sum_{P_{X'|X}}\sum_{P_{X"|X}}2^{-\frac{n}{2}\left(F(P_{XX'})+F(P_{XX"})\right)}\\
&\qquad\times 2^{-n(I_P(X;X')+I_P(X;X")+D(P'_X\|Q)+D(P_{X"}\|Q))} && (537)\\
&= \bigg(\sum_{P_{X'|X}}2^{-\frac{n}{2}F(P_{XX'})}2^{-n\left(I_P(X;X')+D(P'_X\|Q)\right)}\bigg)\\
&\qquad\times\bigg(\sum_{P_{X"|X}}2^{-\frac{n}{2}F(P_{XX"})}2^{-n\left(I_P(X;X")+D(P_{X"}\|Q)\right)}\bigg) && (538)\\
&= \bigg(\sum_{P_{X'|X}}2^{-\frac{n}{2}\left(\sum_{x,x'}d_{\mathrm{B}}(x,x')P_{XX'}(x,x')\right)}\\
&\qquad\times 2^{-n\left(I_P(X;X')+D(P'_X\|Q)\right)}\bigg)^2, && (539)
\end{aligned}
$$

where (537) follows from $\max\{a,b\}\geq\frac{a+b}{2}$. It follows from (539) that

$$
\begin{aligned}
&\mathbb{E}[\mathbb{P}[\{\boldsymbol{X}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{X}_1\to\boldsymbol{X}_3\}]]\\
&= \sum_{\boldsymbol{x}_1}\mathbb{P}(\boldsymbol{x}_1)\mathbb{E}[\mathbb{P}[\{\boldsymbol{x}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{x}_1\to\boldsymbol{X}_3\}]\big|\boldsymbol{X}_1=\boldsymbol{x}_1] && (540)\\
&= \sum_{\boldsymbol{x}_1}\mathbb{P}(\boldsymbol{x}_1)\mathbb{E}[\mathbb{P}[\{\boldsymbol{x}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{x}_1\to\boldsymbol{X}_3\}]] && (541)\\
&= \sum_{P_X}\sum_{\boldsymbol{x}_1\in\mathcal{T}_n(P_X)}\mathbb{P}(\boldsymbol{x}_1)\mathbb{E}[\mathbb{P}[\{\boldsymbol{x}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{x}_1\to\boldsymbol{X}_3\}]] && (542)\\
&= \sum_{P_X}\sum_{\boldsymbol{x}_1\in\mathcal{T}_n(P_X)}2^{-n(D(P_X\|Q)+H(P_X))}\\
&\qquad\times\mathbb{E}[\mathbb{P}[\{\boldsymbol{x}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{x}_1\to\boldsymbol{X}_3\}]] && (543)\\
&\leq \sum_{P_X}\sum_{\boldsymbol{x}_1\in\mathcal{T}_n(P_X)}2^{-n(D(P_X\|Q)+H(P_X))}\\
&\qquad\times\bigg(\sum_{P_{X'|X}}2^{-\frac{n}{2}F(P_{XX'})}2^{-n\left(I_P(X;X')+D(P'_X\|Q)\right)}\bigg)^2 && (544)\\
&\leq \sum_{P_X}2^{-nD(P_X\|Q)}\bigg(\sum_{P_{X'|X}}2^{-\frac{n}{2}F(P_{XX'})}\\
&\qquad\times 2^{-n\left(I_P(X;X')+D(P'_X\|Q)\right)}\bigg)^2, && (545)
\end{aligned}
$$

where (541) follows from the independence of codewords in the random codebook ensemble, while (543) follows from [32, p. 2506].

Now, for all joint types $P_{XX'}$ such that $D(P_{XX'}\|Q_XQ_{X'})>2R$, it holds:

$$
\sum_{n=1}^{\infty}\mathbb{P}\big[\mathcal{N}(P_{XX'})\geq 1\big] \tag{546}
$$

$$\leq \sum_{n=1}^{\infty} \mathbb{E}\big[\mathcal{N}(P_{XX'})\big] \tag{547}$$
$$\leq \sum_{n=1}^{\infty} 2^{-n\left(D(P_{XX'}\|Q_XQ'_X)-2R\right)} < \infty. \tag{548}$$

From (548) and Borel-Cantelli's lemma [24, Theorem 4.3], it holds almost surely that $\mathcal{N}(P_{XX'})=0$ for all joint type $P_{XX'}$ such that $D(Q_{XX'}\|Q_XQ_{X'})>2R$. Hence, from (539) and the above fact with noting the number of types or conditional types are sub-exponential in N, we have

$$\mathbb{E}[\mathbb{P}[\{\boldsymbol{X}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{X}_1\to\boldsymbol{X}_3\}]]$$
$$= 2^{-n\min_{P_{XX'}\in\mathcal{A}_0} D(P_X\|Q)+2\left(I_P(X;X')+D(P'_X\|Q)\right)+F(P_{XX'})}, \tag{549}$$

where (549) follows from the sub-exponential number of possible n-types in $\mathcal{X}\times\mathcal{X}$ [32, p. 2506] Now, note that $Q_X=Q'_X=Q$, so we have

$$I_P(X;X')$$
$$= D(P_{XX'}\|P_XP'_X) \tag{550}$$
$$= D(P_{XX'}\|Q_XQ'_X) - D(P_X\|Q) - D(P'_X\|Q). \tag{551}$$

It follows that

$$D(P_X\|Q)+2\big(I_P(X;X')+D(P'_X\|Q)\big)$$
$$= D(P_X\|Q)+2\big(D(P_{XX'}\|Q_XQ'_X)-D(P_X\|Q)\big) \tag{552}$$
$$= 2D(P_{XX'}\|Q_XQ'_X)-D(P_X\|Q) \tag{553}$$
$$\geq D(P_{XX'}\|Q_XQ'_X), \tag{554}$$

where (554) follows from the data processing for KL divergence (or log-sum inequality [33, Theorem 2.7.1]). Hence, we have

$$\min_{P_{XX'}\in\mathcal{A}_0} D(P_X\|Q)+2\big(I_P(X;X')+D(P'_X\|Q)\big)+F(P_{XX'})$$
$$\geq \min_{P_{XX'}\in\mathcal{A}_0} D(P_{XX'}\|Q_XQ'_X)+F(P_{XX'}) \tag{555}$$
$$= E_{\mathrm{trc}}(R,Q)+R, \tag{556}$$

where (555) follows from (554), and (556) follows from Lemma 7. Note that (555) becomes equality if and only if $P_{XX'}(x,x')=Q(x)Q(x')$ for all $x,x'\in\mathcal{X}\times\mathcal{X}$. However, at $P_{XX'}=Q_XQ'_X$, we have

$$\min_{P_{XX'}\in\mathcal{A}_0} D(P_X\|Q)+F(P_{XX'})$$
$$+2\big(I_P(X;X')+D(P'_X\|Q)\big) \tag{557}$$
$$= \sum_{x,x'} d_{\mathrm{B}}(x,x')Q(x)Q(x') \tag{558}$$
$$= -\sum_{x,x'}\log\left(\sum_{y\in\mathcal{Y}}\sqrt{W(y|x)W(y|x')}\right)Q(x)Q(x') \tag{559}$$
$$> -\log\left(\sum_{x,x'}\sum_{y\in\mathcal{Y}}\sqrt{W(y|x)W(y|x')}Q(x)Q(x')\right) \tag{560}$$
$$= -\log\left(\sum_{y\in\mathcal{Y}}\left(\sum_{x}\sqrt{W(y|x)}Q(x)\right)^2\right) \tag{561}$$
$$= R_0(Q) \tag{562}$$
$$= E_{\mathrm{rce}}(R,Q)+R, \tag{563}$$

where (560) follows from the convexity of the function $-\log x$ noting that the equality does not happen by the condition if $d_{\mathrm{B}}(x,x')$ is not a constant for all (x,x'), and (562) follows from [30, Eq. (8.45)] with $R_0(Q)$ is the cut-off rate of the DMC at the distribution Q. Therefore, from (556) and (563), we have

$$\min_{P_{XX'}\in\mathcal{A}_0} D(P_{XX'}\|Q_XQ'_X)+F(P_{XX'}) > E_{\mathrm{trc}}(R,Q)+R \tag{564}$$

for the case $E^{\mathrm{ub}}_{\mathrm{trc}}(R,Q)=E_{\mathrm{rce}}(R,Q)$.

Now, for the case $E^{\mathrm{ub}}_{\mathrm{trc}}(R,Q)>E_{\mathrm{rce}}(R,Q)$, (556) happens at the optimizer $P^*_{XX'}$ satisfying $D(P^*_{XX'}\|Q_XQ'_X)=2R$, which leads to $P^*_{XX'}\neq Q_XQ'_X$ if $R>0$, so the equality can not happen in (554).

In summary, at $R>0$ and a fixed underlying distribution Q, it holds that

$$\min_{P_{XX'}\in\mathcal{A}_0} D(P_{XX'}\|Q_XQ'_X)+F(P_{XX'}) > E^{\mathrm{ub}}_{\mathrm{trc}}(R,Q)+R. \tag{565}$$

Hence, it holds from (549) and (46) that, for some constant $\delta(R)>0$:

$$\mathbb{E}\Big[\mathbb{P}\big[\{\boldsymbol{X}_1\to\boldsymbol{X}_2\}\cap\{\boldsymbol{X}_1\to\boldsymbol{X}_3\}\Big]$$
$$\leq 2\times 2^{-\left(R+E^{\mathrm{ub}}_{\mathrm{trc}}(R,Q)\right)n}2^{-\delta(R)n}. \tag{566}$$

Now, on the other hand, we know that

$$\mathbb{E}[\mathbb{P}[\boldsymbol{X}_1\to\boldsymbol{X}_2]]$$
$$= \frac{1}{M_n(M_n-1)}\sum_{i=0}^{M_n-1}\sum_{j\neq i}\mathbb{E}[\mathbb{P}[\boldsymbol{X}_i\to\boldsymbol{X}_j]] \tag{567}$$
$$= \frac{\mathbb{E}[P^{\mathrm{ub}}_{\mathrm{e}}(\mathcal{C}_n)]}{M_n-1} \tag{568}$$
$$\geq \frac{\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]}{M_n-1} \tag{569}$$
$$= 2^{-n(E_{\mathrm{rce}}(R,Q)-R)}. \tag{570}$$

From (522), (566), and (570), we obtain (116).

A.9. *Proof of Lemma 11*

We have that

$$\mathbb{E}[\mathcal{I}\{i,j\}\mathcal{I}\{i,k\}]$$
$$= \mathbb{P}\{(X_i,X_j)\in\mathcal{T}_n(P_{XX'}),(X_i,X_k)\in\mathcal{T}_n(P_{XX'})\} \tag{571}$$
$$= \sum_{\boldsymbol{x}_i}\mathbb{P}\{X_i=\boldsymbol{x}_i\}$$
$$\times\mathbb{P}\{(\boldsymbol{x}_i,X_j)\in\mathcal{T}_n(P_{XX'}),(\boldsymbol{x}_i,X_k)\in\mathcal{T}_n(P_{XX'})\} \tag{572}$$
$$= \sum_{\boldsymbol{x}_i}\mathbb{P}\{X_i=\boldsymbol{x}_i\}\mathbb{P}\{(\boldsymbol{x}_i,X_j)\in\mathcal{T}_n(P_{XX'})\}^2 \tag{573}$$
$$= \sum_{P_X}\mathcal{N}(P_X)\mathbb{P}\{\boldsymbol{x}\in\mathcal{T}_n(P_X)\}$$

$$\times \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\}^2 \quad (574)$$

$$\doteq \max_{P_X} \mathcal{N}(P_X) \mathbb{P}\{\boldsymbol{x} \in \mathcal{T}_n(P_X)\}$$
$$\times \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\}^2 \quad (575)$$

where in (572) we conditioned to codeword X_i being equal to a given realization $\boldsymbol{x}_i$, (573) is because X_j and X_k are independent, (574) is because they are also identically distributed, we grouped codewords X_i according to their type P_X and used the fact that $\mathbb{P}\{(\boldsymbol{x}_i, X_j) \in \mathcal{T}_n(P_{XX'})\}$ takes the same value when $\boldsymbol{x}_i$ has the same type. Expression (575) is hard to calculate because of the term $\mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\}$. Therefore we find a lower bound and an upper bound on Eqn. (574). The lower bound is:

$$\sum_{P_X} \mathcal{N}(P_X) \mathbb{P}\{\boldsymbol{x} \in \mathcal{T}_n(P_X)\}$$
$$\times \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\}^2$$
$$\geq \bigg(\sum_{P_X} \mathcal{N}(P_X) \mathbb{P}\{\boldsymbol{x} \in \mathcal{T}_n(P_X)\}$$
$$\times \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\} \bigg)^2 \quad (576)$$
$$= \mathbb{P}\{(X_i, X_j) \in \mathcal{T}_n(P_{XX'})\}^2 \quad (577)$$
$$\doteq 2^{-n2D(P_{XX'} \| Q_X Q'_X)} \quad (578)$$

while the upper bound is:

$$\sum_{P_X} \mathcal{N}(P_X) \mathbb{P}\{\boldsymbol{x} \in \mathcal{T}_n(P_X)\}$$
$$\times \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\}^2$$
$$\leq \sum_{P_X} \mathcal{N}(P_X) \mathbb{P}\{\boldsymbol{x} \in \mathcal{T}_n(P_X)\}$$
$$\times \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\}$$
$$\times \max_{P_X} \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\} \quad (579)$$
$$= \mathbb{P}\{(X_i, X_j) \in \mathcal{T}_n(P_{XX'})\}$$
$$\times \max_{P_X} \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\} \quad (580)$$
$$\doteq 2^{-n[D(P_{XX'} \| Q_X Q'_X) + \eta]} \quad (581)$$

where $\eta = -\frac{1}{n} \log \max_{P_X} \mathbb{P}\{(\boldsymbol{x}, X_j) \in \mathcal{T}_n(P_{XX'}) | \boldsymbol{x} \in \mathcal{T}_n(P_X)\} \leq D(P_{XX'} \| Q_X Q'_X)$, and the inequality follows from (578).

A.10. Proof of Lemma 12

The proof is based on [34, Th. 10]. A similar proof of an equivalent result is presented for the case of constant composition codes in [12]. Notice that, unlike [12] (for constant composition codes), our Lemma 11 gives a bound rather than a dot equality (for i.i.d. codes), which has implications on the minimum exponent starting from which a double exponential decay is found. A full proof of Lemma 11 is available in [35] and is not reported here for a matter of space. The proof is obtained by following a similar approach as in [12, Th. 2], using Lemma 11 to bound the corresponding terms Θ and $\frac{\Delta^2}{8\Theta + 2\Delta}$ in [12, Th. 2].

APPENDIX B

B.1. Proof of Lemma 13

The proof is based on Stein's method in [25, Theorem 3.2]. Let $T = S_n/\sqrt{n}$ and

$$T_i = \frac{1}{\sqrt{n}} \sum_{j \neq i} X_j, \quad \forall i \in [n]. \quad (582)$$

Let f be a bounded function with bounded first and second derivative. Observe that

$$\sqrt{n}\mathbb{E}[Tf(T)]$$
$$= \mathbb{E}\bigg[\sum_{i=1}^n X_i(f(T) - f(T_i) - (T - T_i)f'(T))\bigg]$$
$$+ \mathbb{E}\bigg[\sum_{i=1}^n X_i(T - T_i)f'(T)\bigg] + \mathbb{E}\bigg[\sum_{i=1}^n X_i f(T_i)\bigg]. \quad (583)$$

Now, we have

$$\mathbb{E}[Tf(T) - f'(T)]$$
$$= \frac{1}{\sqrt{n}}\mathbb{E}\bigg[\sum_{i=1}^n X_i\big(f(T) - f(T_i) - (T - T_i)f'(T)\big)\bigg]$$
$$+ \frac{1}{\sqrt{n}}\mathbb{E}\bigg[\sum_{i=1}^n X_i(T - T_i)f'(T)\bigg]$$
$$+ \frac{1}{\sqrt{n}}\mathbb{E}\bigg[\sum_{i=1}^n X_i f(T_i)\bigg] - \mathbb{E}[f'(T)] \quad (584)$$
$$\leq \bigg|\frac{1}{\sqrt{n}}\mathbb{E}\bigg[\sum_{i=1}^n X_i\big(f(T) - f(T_i) - (T - T_i)f'(T)\big)\bigg]\bigg|$$
$$+ \bigg|\frac{1}{\sqrt{n}}\mathbb{E}\bigg[\sum_{i=1}^n X_i f(T_i)\bigg]\bigg|$$
$$+ \bigg|\mathbb{E}\bigg[f'(T)\bigg(1 - \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i(T - T_i)\bigg)\bigg]\bigg| \quad (585)$$
$$\leq \frac{\|f''\|_\infty}{2\sqrt{n}} \sum_{i=1}^n \mathbb{E}|X_i(T - T_i)^2|$$
$$+ \frac{1}{\sqrt{n}}\bigg|\mathbb{E}\bigg[\sum_{i=1}^n X_i f(T_i)\bigg]\bigg|$$
$$+ \frac{\|f'\|_\infty}{n}\mathbb{E}\bigg|\sum_{i=1}^n (1 - X_i^2)\bigg| \quad (586)$$
$$\leq \frac{\|f''\|_\infty}{2n^{3/2}} \sum_{i=1}^n \mathbb{E}|X_i^3| + \frac{1}{\sqrt{n}}\bigg|\mathbb{E}\bigg[\sum_{i=1}^n X_i f(T_i)\bigg]\bigg|$$
$$+ \frac{\|f'\|_\infty}{n}\mathbb{E}\bigg|\sum_{i=1}^n (1 - X_i^2)\bigg|. \quad (587)$$

Now, observe that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}[X_i f(T_i)]$$
$$= \frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times \sum_{i=1}^n x_i f(t_i) \tag{588}$$
$$= \frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times \sum_{i=1}^n x_i f(t_i) \tag{589}$$
$$+ \frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times \sum_{i=1}^n x_i f(t_i) \tag{590}$$
$$= \sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}} \big(1\pm\xi_n\big)\frac{1}{\sqrt{n}}\prod_{i=1}^n \mathbb{P}[X_i=x_i]\sum_{i=1}^n x_i f(t_i) + \frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times \sum_{i=1}^n x_i f(t_i) \tag{591}$$
$$= \sum_{x_1,x_2,\cdots,x_n\in\mathcal{X}^n} \big(1\pm\xi_n\big)\frac{1}{\sqrt{n}}\prod_{i=1}^n \mathbb{P}[X_i=x_i]\sum_{i=1}^n x_i f(t_i) - \sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \big(1\pm\xi_n\big)\frac{1}{\sqrt{n}}\prod_{i=1}^n \mathbb{P}[X_i=x_i]\sum_{i=1}^n x_i f(t_i) + \frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times \sum_{i=1}^n x_i f(t_i) \tag{592}$$
$$= 2\xi_n g_n\|f\|_\infty - \sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \big(1\pm\xi_n\big)\frac{1}{\sqrt{n}}\prod_{i=1}^n \mathbb{P}[X_i=x_i] \times \sum_{i=1}^n x_i f(t_i) + \frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times \sum_{i=1}^n x_i f(t_i), \tag{593}$$

where (593) is because $\mathbb{E}\big[\sum_{i=1}^n X_i f(T_i)\big] = 0$ under the product probability measure $\prod_{i=1}^n \mathbb{P}[X_i=x_i]$ and

$$\frac{\xi_n}{\sqrt{n}}\bigg|\prod_{i=1}^n \mathbb{P}[X_i=x_i]\sum_{i=1}^n x_i f(x_i)\bigg| \le \xi_n\bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^n |x_i|\bigg)\|f\|_\infty \tag{594}$$
$$\le \xi_n g_n \|f\|_\infty. \tag{595}$$

From (593), we obtain

$$\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}[X_i f(T_i)]\bigg| \le \big(1+\xi_n\big)\frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c}\prod_{i=1}^n \mathbb{P}[X_i=x_i]\bigg|\sum_{i=1}^n x_i f(t_i)\bigg| + \frac{1}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,\cdots,X_n=x_n]\bigg|\sum_{i=1}^n x_i f(t_i)\bigg|. \tag{596}$$

From (596), we obtain

$$\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}[X_i f(T_i)]\bigg| \le \big(1+\xi_n\big)\frac{\|f\|_\infty}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c}\prod_{i=1}^n \mathbb{P}[X_i=x_i]\sum_{i=1}^n |x_i| + \frac{\|f\|_\infty}{\sqrt{n}}\sum_{x_1,x_2,\cdots,x_n\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,\cdots,X_n=x_n]\sum_{i=1}^n |x_i| \tag{597}$$
$$\le (1+\xi_n)\|f\|_\infty g_n \mathbb{P}_\Pi[\mathcal{V}^c] + \|f\|_\infty g_n \mathbb{P}[\mathcal{V}^c] \to 0 \tag{598}$$

as $n\to\infty$, where (598) follows from the assumption (222). On the other hand, we have

$$\frac{1}{n}\mathbb{E}\bigg[\bigg|\sum_{i=1}^n (1-X_i^2)\bigg|\bigg] \le (1+\xi_n)\bigg(\frac{1}{n}\bigg)\sum_{x_1,x_2,\cdots,x_n}\prod_{i=1}^n \mathbb{P}(X_i=x_i)\bigg|\sum_{i=1}^n (1-x_i^2)\bigg| + \frac{1}{n}\mathbb{P}[V^c]\sup_{(x_1,x_2,\cdots,x_n)\in\mathcal{V}^c}\bigg|\sum_{i=1}^n (1-x_i^2)\bigg| \tag{599}$$
$$= (1+\xi_n)\bigg(\frac{1}{n}\bigg)\sum_{x_1,x_2,\cdots,x_n}\prod_{i=1}^n \mathbb{P}(X_i=x_i)\bigg|\sum_{i=1}^n (1-x_i^2)\bigg| + \frac{1}{n}\mathbb{P}[V^c]\sup_{(x_1,x_2,\cdots,x_n)\in\mathcal{V}^c}\max\bigg\{\sum_{i=1}^n x_i^2, n\bigg\} \tag{600}$$
$$\le (1+\xi_n)\bigg(\frac{1}{n}\bigg)\sum_{x_1,x_2,\cdots,x_n}\prod_{i=1}^n \mathbb{P}(X_i=x_i)\bigg|\sum_{i=1}^n (1-x_i^2)\bigg| + \mathbb{P}[V^c]g_n \tag{601}$$
$$= (1+\xi_n)\bigg(\frac{1}{n}\bigg)\sum_{x_1,x_2,\cdots,x_n}\prod_{i=1}^n \mathbb{P}(X_i=x_i) \times \bigg|\sum_{i=1}^n (1-x_i^2)\bigg| + o(1) \tag{602}$$
$$\le (1+\xi_n)\bigg(\frac{1}{n}\bigg)\sum_{x_1,x_2,\cdots,x_n}\prod_{i=1}^n \mathbb{P}(X_i=x_i) \times \bigg|\sum_{i=1}^n x_i^2 - \sum_{i=1}^n \mathbb{E}[X_i^2]\bigg| + o(1) \tag{603}$$
$$\le (1+\xi_n)\bigg(\frac{1}{n}\bigg)\sqrt{\mathrm{Var}_\Pi\bigg[\sum_{i=1}^n X_i^2\bigg]} + o(1) \tag{604}$$

$$= (1+\xi_n)\left(\frac{1}{n}\right)\sqrt{\sum_{i=1}^n \mathrm{Var}(X_i^2)} + o(1) \tag{605}$$

$$\leq (1+\xi_n)\sqrt{\frac{\sum_{i=1}^n \mathbb{E}[|X_i|^4]}{n^2}} + o(1), \tag{606}$$

where (602) follows from (222), (603) follows from $\sum_{i=1}^n \mathbb{E}[X_i^2] = n$, (604) follows from Cauchy–Schwarz inequality, and (606) follows from $\mathrm{Var}(X_i^2) \leq \mathbb{E}[|X_i|^4]$. Furthermore, we also have

$$\frac{1}{n^{3/2}}\sum_{i=1}^n \mathbb{E}\big[|X_i|^3\big]$$
$$\leq (1+\xi_n)\left(\frac{1}{n^{3/2}}\right)\sum_{x_1,x_2,\cdots,x_n}\prod_{i=1}^n \mathbb{P}[X_i=x_i]\left(\sum_{i=1}^n |x_i|^3\right) + \mathbb{P}[V^c]\sup_{(x_1,x_2,\cdots,x_n)\in\mathcal{V}^c}\left(\frac{1}{n^{3/2}}\sum_{i=1}^n |x_i|^3\right) \tag{607}$$

$$\leq (1+\xi_n)\left(\frac{1}{n^{3/2}}\right)\sum_{x_1,x_2,\cdots,x_n}\prod_{i=1}^n \mathbb{P}[X_i=x_i]\left(\sum_{i=1}^n |x_i|^3\right) + \mathbb{P}[\mathcal{V}^c]g_n \tag{608}$$

$$\leq (1+\xi_n)\frac{1}{n^{3/2}}\sum_{i=1}^n \mathbb{E}[|X_i|^3] + o(1), \tag{609}$$

where (609) follows from (222). From (587), (598), (602), and (609), we obtain

$$\big|\mathbb{E}[f'(T) - Tf(T)]\big|$$
$$\leq (1+\xi_n)\frac{\|f''\|_\infty}{2n^{3/2}}\sum_{i=1}^n \mathbb{E}[|X_i|^3] + (1+\xi_n)\|f'\|_\infty\sqrt{\frac{\sum_{i=1}^n \mathbb{E}[|X_i|^4]}{n^2}} + o(1) \to 0 \tag{610}$$

as $n\to\infty$ under the conditions (224) and (225). Then, by [25, Th. 3.1], we conclude that $T \xrightarrow{(d)} \mathcal{N}(0,1)$ under the conditions (224) and (225).

Now, observe that

$$\mathrm{Var}(S_n)$$
$$= \mathbb{E}[(X_1+X_2+\cdots+X_n)^2] \tag{611}$$

$$= \sum_{(x_1,x_2,\cdots,x_n)\in\mathcal{V}} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times (x_1+x_2+\cdots+x_n)^2 + \sum_{(x_1,x_2,\cdots,x_n)\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,X_2=x_2,\cdots,X_n=x_n] \times (x_1+x_2+\cdots+x_n)^2 \tag{612}$$

$$\leq (1+\xi_n)\sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{(x_1,x_2,\cdots,x_n)\in\mathcal{V}^c} \mathbb{P}[X_1=x_1,\cdots,X_n=x_n]g_n n \tag{613}$$

$$\leq (1+\xi_n)\sum_{i=1}^n \mathbb{E}[X_i^2] + g_n n\mathbb{P}[\mathcal{V}^c] \tag{614}$$

$$= (1+\xi_n)\sum_{i=1}^n \mathbb{E}[X_i^2] + o(n), \tag{615}$$

where (613) follows from (221), and (615) follows from (222). From (615) and $\xi_n \to 0$ (since $g_n\xi_n \to 0$ and $g_n \geq 1$), it holds that $\mathrm{Var}(S_n) = n + o(n)$. Hence, by Slutsky's theorem [24, p. 334], we finally obtain that

$$\tilde{T} = \frac{S_n}{\sqrt{S_n}} \to \mathcal{N}(0,1). \tag{616}$$

B.2. Proof of Lemma 14

This proof is based on the proof of [25, Prop. 1.2]. Consider the function $h_x(w) = \mathbb{1}\{w \leq x\}$, and the 'smooth' $h_{x,\varepsilon}(w)$ defined to be one for $w \leq x$, zero for $w > x+\varepsilon$, and linear between them. Then, it is clear that $h_{x,\varepsilon} \in \mathcal{V}$ with $a = x$ and $c = \varepsilon$.

First, observe that $\varepsilon h_{x,\varepsilon}(w)$ is 1-Lipschitz and

$$\big\|\varepsilon h_{x,\varepsilon}\big\|_\infty \leq \varepsilon. \tag{617}$$

Hence, it holds that

$$4\sqrt{2\pi}h_{x,4\sqrt{2\pi}} \in \mathcal{H} = \{h \in \mathcal{V} : c \leq 4\sqrt{2\pi}\}, \tag{618}$$

so $\mathcal{H}$ in the definition of Wasserstein metric (cf. Definition 1) is a non-empty set, and $d_W(T,Z)$ is well-defined.

Furthermore, by definition of $d_{W,\mathrm{mod}}(T,Z)$, it holds that

$$d_{W,\mathrm{mod}}(T,Z) \leq \sup_{h\in\mathcal{H}} \mathbb{E}[|h(Z)|] + \mathbb{E}[|h(T)|] \tag{619}$$
$$\leq 2\|h\|_\infty \tag{620}$$
$$= 2c \tag{621}$$
$$\leq 8\sqrt{2\pi}. \tag{622}$$

Now, by setting $\varepsilon = (2\pi)^{1/4}\sqrt{2d_{W,\mathrm{mod}}(T,Z)}$, it holds that

$$\big\|\varepsilon h_{x,\varepsilon}\big\|_\infty \leq (2\pi)^{1/4}\sqrt{2d_{W,\mathrm{mod}}(T,Z)} \tag{623}$$
$$\leq 4\sqrt{2\pi}, \tag{624}$$

where (623) follows from (617), and (624) follows from (622). This means that $\varepsilon h_{x,\varepsilon} \in \mathcal{H}$ since $\varepsilon h_{x,\varepsilon} \in \mathcal{V}$ as mentioned above. Then, we have

$$\mathbb{E}[h_x(T)] - \mathbb{E}[h_x(Z)]$$
$$= \mathbb{E}[h_x(T)] - \mathbb{E}[h_{x,\varepsilon}(Z)] + \mathbb{E}[h_{x,\varepsilon}(Z)] - \mathbb{E}[h_x(Z)] \tag{625}$$
$$\leq \mathbb{E}[h_{x,\varepsilon}(T)] - \mathbb{E}[h_{x,\varepsilon}(Z)] + \mathbb{E}[h_{x,\varepsilon}(Z)] - \mathbb{E}[h_x(Z)] \tag{626}$$
$$= \frac{1}{\varepsilon}\Big(\mathbb{E}[\varepsilon h_{x,\varepsilon}(T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\Big) + \big|\mathbb{E}[h_{x,\varepsilon}(Z)] - \mathbb{E}[h_x(Z)]\big| \tag{627}$$
$$\leq \frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big| + \big|\mathbb{E}[h_{x,\varepsilon}(Z)] - \mathbb{E}[h_x(Z)]\big|. \tag{628}$$

Similarly, by choosing $h_{x,\varepsilon}(\omega)$ to be 1 when $\omega \leq x-\varepsilon$, 0 when $\omega \geq x$, and linear between them, which is also a function in $\mathcal{V}$, we can show that

$$\begin{aligned}
&\mathbb{E}[h_x(Z)] - \mathbb{E}[h_x(T)] \\
&\quad\leq \frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big| \\
&\qquad + \big|\mathbb{E}[h_{x,\varepsilon}(Z)] - \mathbb{E}[h_x(Z)]\big| &(629)\\
&\quad\leq \frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big| \\
&\qquad + \int_x^{x+\varepsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)\big(h_{x,\varepsilon}(z) - h_x(z)\big)dz &(630)\\
&\quad\leq \frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big| + \frac{\varepsilon}{2\sqrt{2\pi}}. &(631)
\end{aligned}$$

From (628) and (629), we obtain

$$\begin{aligned}
&\big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big| \\
&\quad\leq \frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big| + \frac{\varepsilon}{2\sqrt{2\pi}}. &(632)
\end{aligned}$$

Similarly, we also have

$$\begin{aligned}
&\big|\mathbb{P}(-T \leq x) - \mathbb{P}(Z \leq x)\big| \\
&\quad\leq \frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(-T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big| + \frac{\varepsilon}{2\sqrt{2\pi}}. &(633)
\end{aligned}$$

It follows from (632) and (633) that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \min\big\{\big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big|, \big|\mathbb{P}(-T \leq x) - \mathbb{P}(Z \leq x)\big|\big\} \\
&\quad\leq \sup_{h\in\mathcal{H}} \min\left\{\frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big|,\right. \\
&\qquad \left.\frac{1}{\varepsilon}\big|\mathbb{E}[\varepsilon h_{x,\varepsilon}(-T)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\big|\right\} + \frac{\varepsilon}{2\sqrt{2\pi}} &(634)\\
&\quad= \frac{1}{\varepsilon} d_{W,\mathrm{mod}}(T,Z) + \frac{\varepsilon}{2\sqrt{2\pi}} &(635)\\
&\quad= (8\pi)^{-1/4}\sqrt{d_{W,\mathrm{mod}}(T,Z)}, &(636)
\end{aligned}$$

where (635) follows from $\varepsilon h_{x,\varepsilon} \in \mathcal{H}$, and (636) follows from our setting $\varepsilon = (2\pi)^{1/4}\sqrt{2 d_{W,\mathrm{mod}}(T,Z)}$ above.

Now, for any $x \in \mathbb{R}$, we have

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \min\left\{\big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big|,\right. \\
&\qquad \left.\big|\mathbb{P}(T \leq -x) - \mathbb{P}(Z \leq x)\big|\right\} \\
&\quad\geq \min\left\{\big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big|,\right. \\
&\qquad \left.\big|\mathbb{P}(T \leq -x) - \mathbb{P}(Z \leq x)\big|\right\} &(637)\\
&\quad\geq \min\left\{\big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big|,\right. \\
&\qquad \left.\big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big| - \big|\mathbb{P}(T \leq x) - \mathbb{P}(T \geq -x)\big|\right\} &(638)\\
&\quad\geq \big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big| - \big|\mathbb{P}(T \leq x) - \mathbb{P}(T \geq -x)\big|, &(639)
\end{aligned}$$

where (638) follows from the triangle inequality. From (636) and (639), we obtain (265).

Now, if the distribution of T is tight, then there exists a distribution $\tilde{Y}$ such that $T \xrightarrow{(\mathrm{d})} \tilde{Y}$ [24, p. 337]. Then, if x is a continuous point of $\mathbb{P}(\tilde{Y} \leq x)$ such that $x \to 0$ as $n \to \infty$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \min\left\{\big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big|,\right. \\
&\qquad \left.\big|\mathbb{P}(T \leq -x) - \mathbb{P}(Z \leq x)\big|\right\} \\
&\quad= \lim_{n\to\infty} \min\left\{\big|\mathbb{P}(\tilde{Y} \leq x) - \mathbb{P}(Z \leq x)\big|,\right. \\
&\qquad \left.\big|\mathbb{P}(\tilde{Y} \leq -x) - \mathbb{P}(Z \leq x)\big|\right\} &(640)\\
&\quad= \min\left\{\big|\mathbb{P}(\tilde{Y} \leq 0) - \mathbb{P}(Z \leq 0)\big|,\right. \\
&\qquad \left.\big|\mathbb{P}(\tilde{Y} \leq 0) - \mathbb{P}(Z \leq 0)\big|\right\} &(641)\\
&\quad= \lim_{n\to\infty} \big|\mathbb{P}(T \leq x) - \mathbb{P}(Z \leq x)\big|, &(642)
\end{aligned}$$

where (640) follows from $\lim_{n\to\infty} \min\{A_n, B_n\} = \min\{\lim_{n\to\infty} A_n, \lim_{n\to\infty} B_n\}$ if both the limits $\lim_{n\to\infty} A_n$ and $\lim_{n\to\infty} B_n$ exist. Hence, we obtain (266) from (642) and (636).

B.3. Proof of Lemma 15

By definition 1, we have

$$\begin{aligned}
d_{W,\mathrm{mod}}(T,Z) &= \sup_{h\in\mathcal{H}} \min\left\{\big|\mathbb{E}[h(T)] - \mathbb{E}[h(Z)]\big|,\right. \\
&\qquad \left.\big|\mathbb{E}[h(-T)] - \mathbb{E}[h(Z)]\big|\right\}. &(643)
\end{aligned}$$

Using (267), we obtain the following equalities

$$\begin{aligned}
\mathbb{E}[h(T)] - \mathbb{E}[h(Z)] &= \mathbb{E}\big[f_h'(T) - T f_h(T)\big] &(644)\\
\mathbb{E}[h(-T)] - \mathbb{E}[h(Z)] &= \mathbb{E}\big[f_h'(-T) + T f_h(-T)\big]. &(645)
\end{aligned}$$

Combining (643), (644), and (645), we obtain (268).

B.4. Proof of Lemma 16

By Lemma 15, we have

$$\begin{aligned}
d_{W,\mathrm{mod}}(T,Z) &= \sup_{h\in\mathcal{H}} \min\left\{\big|\mathbb{E}\big[f_h'(T) - T f_h(T)\big]\big|,\right. \\
&\qquad \left.\big|\mathbb{E}\big[f_h'(-T) + T f_h(-T)\big]\big|\right\}. &(646)
\end{aligned}$$

Now, observe that

$$\begin{aligned}
&\mathbb{E}\big[f_h'(T) - T f_h(T)\big] \\
&\quad= \mathbb{E}[h(T)] - \mathbb{E}[h(Z)] &(647)\\
&\quad= \mathbb{E}[h(T_1)] - \mathbb{E}[h(Z)] + \mathbb{E}[h(T) - h(T_1)] &(648)\\
&\quad= \mathbb{E}[f_h'(T_1) - T_1 f_h(T_1)] + \mathbb{E}[h(T) - h(T_1)], &(649)
\end{aligned}$$

where (647) and (649) follow from (267). It follows that

$$\left|\mathbb{E}\big[f_h'(T) - Tf_h(T)\big]\right|$$
$$= \left|\mathbb{E}[f_h'(T_1) - T_1 f_h(T_1)] + \mathbb{E}[h(T) - h(T_1)]\right| \quad (650)$$
$$\leq \left|\mathbb{E}[f_h'(T_1) - T_1 f_h(T_1)]\right| + \left|\mathbb{E}[h(T) - h(T_1)]\right| \quad (651)$$
$$= \left|\mathbb{E}[f_h'(T_1) - T_1 f_h(T_1)]\right| + \mathbb{E}[h(T) - h(T_1)] \quad (652)$$

where (652) follows from $T \leq T_1$ and h is non-increasing. Similarly, we have

$$\left|\mathbb{E}\big[f_h'(-T) + Tf_h(-T)\big]\right|$$
$$= \left|\mathbb{E}[f_h'(-T_1) + T_1 f_h(-T_1)] + \mathbb{E}[h(-T) - h(-T_1)]\right| \quad (653)$$
$$\leq \left|\mathbb{E}[f_h'(-T_1) + T_1 f_h(-T_1)]\right| + \mathbb{E}[h(-T_1) - h(-T)], \quad (654)$$

where (654) follows from $T \leq T_1$ and h is non-increasing. From (652) and (654), for all $h \in \mathcal{H}$, we have

$$\min\left\{\left|\mathbb{E}\big[f_h'(T) - Tf_h(T)\big]\right|, \left|\mathbb{E}\big[f_h'(-T) + Tf_h(-T)\big]\right|\right\}$$
$$\leq \max\left\{\left|\mathbb{E}[f_h'(T_1) - T_1 f_h(T_1)]\right|, \left|\mathbb{E}[f_h'(T_1) - T_1 f_h(T_1)]\right|\right\}$$
$$+ \min\left\{\mathbb{E}[h(T) - h(T_1)], \mathbb{E}[h(-T_1) - h(-T)]\right\}, \quad (655)$$

where (655) follows from $\min\{a+c, b+d\} \leq \max\{a,b\} + \min\{c,d\}$ for all $a,b,c,d \in \mathbb{R}$.

Finally, we obtain (269) from (655).

Appendix C

C.1. Proof of Lemma 19

The proof of this lemma is based on the proof of the [25, Th. 3.1]. Given $h \in \mathcal{H}$, we choose f_h be a solution of the following ODE equation:

$$f_h'(w) - wf_h(w) = h(w) - \Phi(h) \quad (656)$$

where $\Phi(h) = \mathbb{E}[h(Z)]$ with $Z \sim \mathcal{N}(0,1)$, then we have

$$f_h(w) = e^{\frac{w^2}{2}} \int_w^\infty e^{-\frac{t^2}{2}}\big(\Phi(h) - h(t)\big)dt \quad (657)$$
$$= -e^{\frac{w^2}{2}} \int_{-\infty}^w e^{-\frac{t^2}{2}}\big(\Phi(h) - h(t)\big)dt. \quad (658)$$

Now, from (658) the facts listed below follow (see [25, Lemma 2.5]):

$$\|f_h\|_\infty \leq 2\|h'\|_\infty = 2, \quad (659)$$
$$\|f_h^{'}\|_\infty \leq \sqrt{\frac{2}{\pi}}\|h'\|_\infty = \sqrt{\frac{2}{\pi}}, \quad (660)$$
$$\|f_h^{"}\|_\infty \leq 2\|h'\|_\infty = 2. \quad (661)$$

Now, assume that $\mathbb{E}[T] = 0$ and $\mathbb{E}[T^2 = 1]$. Furthermore, for any $h \in \mathcal{H}$, from (656), it holds that

$$|f_h^{'}(T) - Tf_h(T)| = |h(T) - \Phi(h)| \quad (662)$$
$$= |h(T) - \mathbb{E}[h(Z)]| \quad (663)$$
$$\leq 2\|h\|_\infty \quad (664)$$
$$\leq 8\sqrt{2\pi}. \quad (665)$$

Furthermore, from (656), we also have

$$\tilde{d}_{W,\mathrm{mod}}(T,Z) = \sup_{h \in \mathcal{H}} \big|\mathbb{E}[h(T)] - \mathbb{E}[h(Z)]\big| \quad (666)$$
$$\leq \sup_{f_h : h \in \mathcal{H}} \big|\mathbb{E}[Tf_h(T) - f_h^{'}(T)]\big|. \quad (667)$$

Now, for all $f_h : h \in \mathcal{H}$, observe that

$$Tf_h(T) - f_h^{'}(T)$$
$$= T\big(f_h(T) - f_h(0) - Tf_h^{'}(0)\big) + Tf_h(0)$$
$$+ \big(T^2 - 1\big)f_h^{'}(0) + \big(f_h^{'}(0) - f_h^{'}(T)\big). \quad (668)$$

It follows from (668) that

$$\mathbb{E}\big[Tf_h(T) - f_h^{'}(T)\big]$$
$$= \mathbb{E}\big[T\big(f_h(T) - f_h(0) - Tf_h^{'}(0)\big)\big] + f_h(0)\mathbb{E}[T]$$
$$+ f_h^{'}(0)\mathbb{E}\big[T^2 - 1\big] + \mathbb{E}[f_h^{'}(0) - f_h^{'}(T)] \quad (669)$$
$$= \mathbb{E}\big[T\big(f_h(T) - f_h(0) - Tf_h^{'}(0)\big)\big] + \mathbb{E}[f_h^{'}(0) - f_h^{'}(T)], \quad (670)$$

where (670) follows from the fact that $\mathbb{E}[T] = 0$ and $\mathbb{E}[T^2] = 1$. Hence, from (667) and (670), we have

$$\tilde{d}_{W,\mathrm{mod}}(T,Z)$$
$$\leq \sup_{f_h : h \in \mathcal{H}} \big|\mathbb{E}[Tf_h(T) - f_h^{'}(T)]\big| \quad (671)$$
$$\leq \sup_{f_h : h \in \mathcal{H}} \mathbb{E}\big[\big|T\big(f_h(T) - f_h(0) - Tf_h^{'}(0)\big)\big|\big]$$
$$+ \mathbb{E}[\big|f_h^{'}(0) - f_h^{'}(T)\big|]. \quad (672)$$

Now, observe that

$$\big|T\big(f_h(T) - f_h(0) - Tf_h^{'}(0)\big)\big|$$
$$= \big|Tf_h(T) - f_h^{'}(T) + f_h^{'}(T) - Tf_h(0) - T^2 f_h^{'}(0)\big| \quad (673)$$
$$= \big|Tf_h(T) - f_h^{'}(T) + f_h^{'}(T)$$
$$- Tf_h(0) - f_h^{'}(0) + (1 - T^2)f_h^{'}(0)\big| \quad (674)$$
$$\leq \big|Tf_h(T) - f_h^{'}(T)\big| + |f_h^{'}(T)| + |Tf_h(0)|$$
$$+ |f_h^{'}(0)| + |f_h^{'}(0)(T^2 - 1)|\big) \quad (675)$$
$$\leq 8\sqrt{2\pi} + 2\sqrt{\frac{2}{\pi}} + 2|T| + \sqrt{\frac{2}{\pi}}|T^2 - 1| \quad (676)$$
$$= \left(8 + \frac{2}{\pi}\right)\sqrt{2\pi} + 2|T| + \sqrt{\frac{2}{\pi}}|T^2 - 1|, \quad (677)$$

where (676) follows from (659), (660) and (665). Hence, we have

$$\begin{aligned}&\big|T\big(f_h(T)-f_h(0)-Tf_h^{'}(0)\big)\big|\\&\quad=\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi}+2|T|,\right.\\&\qquad\left.\big|T\big(f_h(T)-f_h(0)-Tf_h^{'}(0)\big)\big|\right\} &(678)\\&\quad\le\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},\big|T\big(f_h(T)-f_h(0)-Tf_h^{'}(0)\big)\big|\right\}\\&\qquad+2|T|+\sqrt{\frac{2}{\pi}}|T^2-1|, &(679)\end{aligned}$$

where (679) follows from $\min\{A+B,C\}\le\min\{A,C\}+B$ for all $A,B,C\ge 0$. It follows from (679) that

$$\begin{aligned}&\mathbb{E}\big[\big|T\big(f_h(T)-f_h(0)-Tf_h^{'}(0)\big)\big|\big]\\&\quad\le\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},\right.\right.\\&\qquad\left.\left.\big|T\big(f_h(T)-f_h(0)-Tf_h^{'}(0)\big)\big|\right\}\right]+2\mathbb{E}\big[|T|\big]. &(680)\end{aligned}$$

Now, by Taylor's expansion, for some $\eta\in(0,-|T|)\cup(0,|T|)$, we have

$$\begin{aligned}\big|T\big(f_h(T)-f_h(0)-Tf_h^{'}(0)\big)\big|&=\frac{1}{2}\big|T^3f_h^{"}(\eta)\big| &(681)\\&\le\frac{1}{2}\|f_h^{"}\|_\infty|T^3| &(682)\\&\le|T^3|. &(683)\end{aligned}$$

Hence, from (680) and (683), we obtain

$$\begin{aligned}&\mathbb{E}\big[\big|T\big(f_h(T)-f_h(0)-Tf_h^{'}(0)\big)\big|\big]\\&\quad\le\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|^3\right\}\right]\\&\qquad+2\mathbb{E}\big[|T|\big]+\sqrt{\frac{2}{\pi}}|T^2-1|. &(684)\end{aligned}$$

Similarly, by Taylor's expansion, for some $\theta\in(0,-|T|)\cup(0,|T|)$, we obtain

$$\begin{aligned}\mathbb{E}\big[\big|f_h^{'}(T)-f_h^{'}(0)\big|\big]&=\mathbb{E}\big[\big|f_h^{"}(\theta)T\big|\big] &(685)\\&\le\mathbb{E}\big[|f_h^{"}(\theta)\|T|\big] &(686)\\&\le\|f_h^{"}\|_\infty\mathbb{E}\big[|T|\big] &(687)\\&\le 2\mathbb{E}[|T|]. &(688)\end{aligned}$$

Finally, from (679), (684), and (688), we have

$$\begin{aligned}&\tilde{d}_{W,\mathrm{mod}}(T,Z)\\&\quad\le\sup_{f_h:h\in\mathcal{H}}\mathbb{E}\big[\big|T\big(f_h(T)-f_h(0)-Tf_h'(0)\big)\big|\big]\\&\qquad+\mathbb{E}\big[\big|f_h^{'}(0)-f_h^{'}(T)\big|\big] &(689)\\&\quad\le\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|^3\right\}\right]\\&\qquad+2\mathbb{E}\big[|T|\big]+\sqrt{\frac{2}{\pi}}|T^2-1|. &(690)\end{aligned}$$

Now, observe that

$$\begin{aligned}&\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|^3\right\}\right]\\&=\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|^3\right\}\middle\||T|\le 1\right]\mathbb{P}[|T|\le 1]\\&\quad+\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|^3\right\}\middle\||T|>1\right]\mathbb{P}[|T|>1] &(691)\\&\le\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|\right\}\middle\||T|\le 1\right]\mathbb{P}[|T|\le 1]\\&\quad+\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|\right\}\middle\||T|>1\right]\mathbb{P}[|T|>1] &(692)\\&\le\mathbb{E}\left[|T|\middle\||T|\le 1\right]\mathbb{P}[|T|\le 1]\\&\qquad+\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|^3\right\}\middle\||T|>1\right]\\&\qquad\times\mathbb{P}[|T|>1] &(693)\\&\le\mathbb{E}\big[|T|\big]+\mathbb{E}\left[\min\left\{\left(8+\frac{2}{\pi}\right)\sqrt{2\pi},|T|^3\right\}\middle\||T|>1\right]\\&\qquad\times\mathbb{P}[|T|>1] &(694)\\&\le\mathbb{E}\big[|T|\big]+\left(8+\frac{2}{\pi}\right)\sqrt{2\pi}\mathbb{P}[|T|>1] &(695)\\&\le\mathbb{E}\big[|T|\big]+\left(8+\frac{2}{\pi}\right)\sqrt{2\pi}\mathbb{E}[|T|] &(696)\\&\le\left(10+\frac{1}{\pi}\right)\sqrt{2\pi}\mathbb{E}[|T|], &(697)\end{aligned}$$

where (692) follows from $|T|^3\le|T|$ for all $|T|\le 1$, (694) follows from $\mathbb{E}[X]=\mathbb{E}[X|A]\mathbb{P}(A)+\mathbb{E}[X|A^c]\mathbb{P}(A^c)\ge\mathbb{E}[X|A]\mathbb{P}(A)$ for all non-negative random variable X, and (696) follows from Markov's inequality.

From (690) it follows that

$$\begin{aligned}&\tilde{d}_{W,\mathrm{mod}}(T,Z)\\&\quad\le\left(\left(10+\frac{1}{\pi}\right)\sqrt{2\pi}+\left(4+\sqrt{\frac{2}{\pi}}\right)\right)\mathbb{E}[|T|]\\&\qquad+\sqrt{\frac{2}{\pi}}\mathbb{E}\big[|T^2-1|\big] &(698)\\&\quad<40\mathbb{E}[|T|]+\sqrt{\frac{2}{\pi}}\mathbb{E}\big[|T^2-1|\big]. &(699)\end{aligned}$$

By combining Lemma 14 and (699), we have

$$\begin{aligned}d_K(T,Z)&<2(8\pi)^{-1/4}\sqrt{40\mathbb{E}[|T|]+\sqrt{\frac{2}{\pi}}\mathbb{E}\big[|T^2-1|\big]} &(700)\\&\le 14(8\pi)^{-1/4}\sqrt{\mathbb{E}[|T|]+\mathbb{E}\big[|T^2-1|\big]}. &(701)\end{aligned}$$

This concludes the proof.

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," 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. Nakiboğlu, "The sphere packing bound for memoryless channels," *Problems of Information Transmission*, vol. 56, pp. 201–244, 2020.

[6] ——, "The Augustin capacity and center," *Problems of Information Transmission*, vol. 55, no. 4, pp. 299–342, 2019.

[7] A. Barg and G. D. Forney, "Random codes: Minimum distances and error exponents," vol. 48, no. 9, pp. 2568–2573, 2002.

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

[9] N. Merhav, "Error exponents of typical random codes," vol. 64, no. 9, pp. 6223–6235, 2018.

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

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

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

[13] 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.

[14] R. Tamir and N. Merhav, "Universal decoding for the typical random code and for the expurgated code," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2156–2168, 2022.

[15] G. Cocco, A. Guillén i Fàbregas, and J. Font-Segura, "A dual-domain achievability of the typical error exponent," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 776–793, Feb. 2023.

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

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

[18] G. Poltyrev, "Random coding bounds for discrete memoryless channels," *Prob. Inf. Transm.*, vol. 18, no. 1, pp. 9–21, 1982.

[19] R. G. Gallager. (1994) Fixed composition arguments and lower bounds to the error probability, lecture notes MIT 6.441. [Online]. Available: https://www.rle.mit.edu/rgallager/documents/notes5.pdf,tinyurl.com/22akp3ba

[20] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of a discrete memoryless channel," in *Proc. ISIT*, vol. 77, 1977, p. 107.

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

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

[23] M. Veraar, "On Khintchine inequalities with a weight," *Proceedings of the American Mathematical Society*, vol. 138, pp. 4119–4121, 2010.

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

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

[26] D. de Caen, "A lower bound on the probability of a union," *Discrete Mathematics*, vol. 169, no. 1, pp. 217–220, 1997.

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

[28] H. Royden and P. Fitzpatrick, *Real Analysis*, 4th ed. Pearson, 2010.

[29] R. Gallager, "The random coding bound is tight for the average code (corresp.)," vol. 19, no. 2, pp. 244–246, 1973.

[30] S. M. Moser, *Advanced Topics in Information Theory. Lecture Notes*, 2019.

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

[32] I. Csiszár, "The method of types," *IEEE Trans. Inf. Th.*, vol. 44, no. 6, pp. 2505–23, 1998.

[33] T. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Th.*, vol. 44, no. 6, pp. 2524–30, 1998.

[34] S. Janson, "New versions of Suen's correlation inequality," *Random Struct. Algorithms*, vol. 13, no. 3-4, pp. 467–483, Oct. 1998.

[35] L. V. Truong, G. Cocco, J. Font-Segura, and A. Guillén i Fàbregas, "Concentration properties of random codes," 2022. [Online]. Available: https://arxiv.org/abs/2203.07853

Lan V. Truong (S'12-M'15) received the B.S.E. degree in Electronics and Telecommunications from the Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam, in 2003, and the M.S.E. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, in 2011, and the Ph.D. degree from the Department of Electrical and Computer Engineering, National University of Singapore (NUS), Singapore, in 2018. He was an Operation and Maintenance Engineer with MobiFone Telecommunications Corporation, Hanoi, for several years. He spent one year as a Research Assistant with the NSF Center for Science of Information and the Department of Computer Science, Purdue University, in 2012. From 2013 to 2015, he was a University Lecturer with the Department of Information Technology Specialization, FPT University, Hanoi, Vietnam. From 2018 to 2019, he was a Research Fellow with the Department of Computer Science, School of Computing, NUS. From 2020 to 2023, he was a Research Associate with the Department of Engineering, University of Cambridge, United Kingdom. Presently, he is a Lecturer (Assistant Professor) at the School of Mathematics, Statistics and Actuarial Science, University of Essex, United Kingdom. His research interests include information theory, machine learning, data science, and probability.

Giuseppe Cocco (Senior Member, IEEE) received the M.S. degree (Hons.) in Telecommunications Engineering from Università di Pisa, Pisa, Italy, in 2008, and the Ph.D. degree in Telecommunications Engineering from the Universitat Politècnica de Catalonia (UPC), Barcelona, Spain, in 2013.

In 2009, he was awarded the Networking/Partnering Initiative grant (Ph.D. co-funding) by the European Space Agency (ESA). His Ph.D. studies were co-funded by ESA and the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona. From 2010 to 2011, he spent one year as Visiting Ph.D. Candidate at the European Space Research and Technology Centre (ESTEC), ESA, Noordwijk, The Netherlands. From 2014 to 2017, he was with the Institute of Communications and Navigation, German Aerospace Center (DLR), Weßling, Germany, where he was Senior Researcher and Project Manager. In 2017, he was awarded a Marie Skłodowska Curie Individual fellowship grant by the European Commission and from 2017 to 2019 he was at the Laboratory of Intelligent Systems (robotics) and the LTS4 Signal Processing Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. In 2019, he was awarded the Beatriu de Pinós fellowship by the Catalan Government and joined the Information Theory and Coding group at the Universitat Pompeu Fabra (UPF), Barcelona.

Since 2023 he is Ramon y Cajal Assistant Professor at the UPC in Barcelona, Spain. His research interests include Information Theory, Quantum Information and satellite communications.

Josep Font-Segura (Member, IEEE) received the PhD degree in electrical engineering from Universitat Politècnica de Catalunya, Barcelona, in 2014.

He was graduate research assistant at Columbia University, New York, in 2008–2009, doctoral research assistant at Universitat Politècnica de Catalunya, Barcelona, in 2009–2014 and postdoctoral research assistant at Universitat Pompeu Fabra, Barcelona, in 2014–2023. Since 2023, he is assistant professor at Universitat Pompeu Fabra. His research interests are in the areas of information theory, mathematical physics, communications, and signal processing.

Dr Font-Segura is a member of the IEEE. He is recipient of a "la Caixa" fellowship for graduate studies in the United States, a best paper award at the IEEE International Conference on Communications, and a Juan de la Cierva postdoc fellowship from the Spanish government. He was member of the organizing committee of the IEEE International Symposium on Information Theory in 2016, and he was vice-chair of the Spanish chapter of the IEEE Information Theory Society in 2018-2019. He also contributed as reviewer in several journal and conference publications of the IEEE society.

Albert Guillén i Fàbregas (Fellow, IEEE) received the Telecommunications Engineering Degree and the Electronics Engineering Degree from Universitat Politècnica de Catalunya and Politecnico di Torino, respectively in 1999, and the Ph.D. in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL) in 2004.

He is a Professor at the Department of Engineering, University of Cambridge and a part-time researcher at Universitat Pompeu Fabra. He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, Universitat Pompeu Fabra, University of Cambridge, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra (ICREA Research Professor), University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His specific research interests are in the areas of information theory, communication theory, coding theory, statistical inference.

Dr. Guillén i Fàbregas is a Member of the Young Academy of Europe, and received the Starting and Consolidator Grants from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. He is a Fellow of the IEEE and of the Institute for Mathematics and its Applications (IMA). Since 2013 he has been an Editor of the Foundations and Trends in Communications and Information Theory, Now Publishers and was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY (2013–2020) AND IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2011).