Generalized Random Gilbert-Varshamov Codes: Typical Error Exponent and Concentration Properties

Lan V. Truong, *Member, IEEE*, and Albert Guillén i Fàbregas, *Fellow, IEEE*

***Abstract*— We find the exact typical error exponent of constant composition generalized random Gilbert-Varshamov (RGV) codes over discrete memoryless channels with generalized likelihood decoding. We show that the typical error exponent of the RGV ensemble is equal to the expurgated error exponent, provided that the RGV codebook parameters are chosen appropriately. We also prove that the random coding exponent converges in probability to the typical error exponent, and the corresponding non-asymptotic concentration rates are derived. Our results show that the decay rate of the lower tail is exponential while that of the upper tail is double exponential above the expurgated error exponent. The explicit dependence of the decay rates on the RGV distance functions is characterized.**

***Index Terms*— Random coding, error exponent, typical error exponent, gilbert-Varshamov codes, concentration properties.**

I. Introduction

INTRODUCED by Shannon [1], random coding is the key technique employed in information theory in order to show that a code with low error probability exists without explicitly constructing it. Codes are constructed at random, and the average error probability over all randomly generated codes is bounded. Then, it follows that there must exist a code with error probability at least as low as the ensemble average error probability over the codes. In particular, for discrete memoryless channel (DMC), Shannon showed that there exists a code of rate smaller than the channel capacity with vanishing probability of error as the codeword length increases.

Since Shannon's work, random coding has not only been applied extensively, but has been refined in a number of ways. For rates below capacity, Fano [2] characterized the exponential decay of the error probability defining the random coding exponent (RCE) as the negative normalized logarithm of the ensemble-average error probability. In [3], Gallager derived the RCE in a simpler way and introduced the idea of expurgation in order to show the existence of a code with an improved exponent the at low rates. An upper bound to the error exponent for 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].

Most proofs invoking random coding arguments, assume that codewords are independent. Random Gilbert-Varshamov (RGV) codes were first introduced in [7], and are a family of random codes inspired by the basic construction attaining the Gilbert-Varshamov bound for codes in Hamming spaces. The code construction is based on drawing codewords recursively from a fixed type class, in such a way that a newly generated codeword must be at a certain minimum distance from all previously chosen codewords, according to some generic distance function. For suitably optimized parameters, the RCE of RGV codes with maximum-likelihood (ML) decoding is the Csiszár and Körner's exponent [8], which is known to be at least as high as both the random-coding and expurgated exponents.

Most works on random coding and error exponents study the RCE, the error exponent of the ensemble-average error probability. In [9], Barg and Forney studied i.i.d. random coding over the binary symmetric channel (BSC) with ML and showed that the error exponent of most random codes is close to the so-called typical random coding (TRC) exponent, strictly higher than the RCE at low rates. Upper and lower bounds on the TRC for constant-composition codes and general DMCs were provided in [10]. For the same type of codes and channels, Merhav [11] determined the exact TRC error exponent for a generic stochastic decoder called generalized likelihood decoder (GLD), of which ML is a special case. Merhav derived the TRC exponent for spherical codes over coloured Gaussian channels [12] and for random convolutional code ensembles [13], and provided a dual expression of the TRC for i.i.d. codes in [14]. Tamir et al. [15] studied the upper and lower tails of the error exponent around the TRC 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 doubly-exponentially; the latter was first

Manuscript received 22 November 2022; revised 18 June 2023; accepted 26 August 2023. Date of publication 30 August 2023; date of current version 22 January 2024. This work was supported in part by the European Research Council under European Research Council (ERC) under Grant 725411 and in part by the Spanish Ministry of Economy and Competitiveness under Grant PID2020-116683GB-C22. An earlier version of this paper was presented in part at the 2023 IEEE Information Theory Workshop [DOI: 10.1109/ITW55543.2023.10161687]. *(Corresponding author: Lan V. Truong.)*

Lan V. Truong is with the School of Mathematics, Statistics and Actuarial Science, University of Essex, CO4 3SQ Colchester, U.K. (e-mail: lan.truong@essex.ac.uk).

Albert Guillén i Fàbregas is with the Department of Engineering, University of Cambridge, CB2 1PZ Cambridge, U.K., and also with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: guillen@ieee.org).

Communicated by A. Anastasopoulos, Associate Editor for Communications.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3310117.

Digital Object Identifier 10.1109/TIT.2023.3310117

established for a limited range of rates in [16]. By studying the behavior of both tails, work in [15] proves concentration in probability. The TRC was shown to be universally achievable with a likelihood mutual-information decoder in [17]. For pairwise-independent ensembles and arbitrary channels, Cocco et al. showed in [18] that the probability that a code in the ensemble has an exponent smaller than a lower bound on the TRC exponent is vanishingly small. Recently, Truong et al. showed that, for DMCs, the error exponent of a randomly generated code with pairwise-independent codewords converges in probability to its expectation – the typical error exponent [19]. For high rates, the result is a consequence of the fact that the RCE 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 probability of error. Paper [19] also zooms into the behavior at asymptotically low rates and shows that the error exponent converges in distribution to Gaussian-like distributions. From this body of works it emerges that the TRC is the fundamental error exponent attained by specific random-coding ensembles. The performance of poor codes has a critical role in the RCE, while it does not count much towards the TRC.

A. Contributions

This work focusses on the RGV code ensemble and discusses concentration properties of error exponents around its TRC. Compared with constant-composition codes, the dependence among RGV codewords causes standard concentration inequalities such as Hoeffding's inequality not to hold. In this work, we develop new techniques to overcome the challenges presented by RGV codeword dependence. Our main contributions include:

- We find the exact TRC for the RGV ensemble by proving matching upper and lower bounds on the TRC and show that it is equal to Merhav's expurgated exponent [20] for suitably optimized distance function and minimum distance. In addition, we show that for ML decoding, the TRC of the RGV ensemble is at least as high as the maximum of the expurgated exponent and RCE for constant composition codes.
- We show that the random error exponent converges in probability to the TRC.
- We characterize the convergence rates of the above convergence and show that it is exponential for the lower tail and double-exponential for the upper tail under some technical conditions.

B. Notation

Random variables will be denoted by capital letters, and their realizations will be denoted by the corresponding lower case letters. Random vectors and their realizations will be denoted, respectively, by boldfaced capital and lower case letters. Their alphabets will be superscripted by their dimensions. For a generic joint distribution $P_{XY} = \{P_{XY}(x,y), x \in \mathcal{X}, y \in \mathcal{Y}\}$, which will often be abbreviated by P, information measures will be denoted in the conventional manner, but with a subscript P, that is $I_P(X;Y)$ is the mutual information between X and Y, and similarly for other quantities. Natural logarithms are assumed in the derivations; examples will employ base 2. The probability of an event $\mathcal{E}$ will be denoted by $\mathbb{P}[\mathcal{E}]$, the indicator function of event $\mathcal{E}$ will be denoted by $\mathbb{1}\{\mathcal{E}\}$, and the expectation operator will be denoted by $\mathbb{E}[\cdot]$. The notation $[t]_+$ will stand for $\max\{t, 0\}$.

For two positive sequences, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ will stand for exponential equality, that is $\lim_{n\to\infty} \frac{1}{n}\log(\frac{a_n}{b_n}) = 0$. Exponential inequalities $a_n \dot{\leq} b_n$ and $a_n \dot{\geq} b_n$ are defined as $\lim_{n\to\infty} \frac{1}{n}\log(\frac{a_n}{b_n}) \leq 0$ and $\lim_{n\to\infty} \frac{1}{n}\log(\frac{a_n}{b_n}) \geq 0$, respectively. Accordingly, the notation $a_n \doteq e^{-n\infty}$ means that a_n decays super-exponentialy. For two positive sequences, $\{a_n\}$ and $\{b_n\}$, whose elements are both smaller than one for all large enough n, the notation $a_n \stackrel{\circ}{=} b_n$ will stand for double-exponential equality, that is

$$\lim_{n\to\infty} \frac{1}{n}\log\left(\frac{\log b_n}{\log a_n}\right) = 0. \tag{1}$$

Similarly, $a_n \stackrel{\circ}{\leq} b_n$ means that

$$\limsup_{n\to\infty} \frac{1}{n}\log\left(\frac{\log b_n}{\log a_n}\right) \leq 0, \tag{2}$$

and $a_n \stackrel{\circ}{\geq} b_n$ stands for

$$\liminf_{n\to\infty} \frac{1}{n}\log\left(\frac{\log b_n}{\log a_n}\right) \geq 0. \tag{3}$$

A sequence of random variables $\{A_n\}_{n=1}^{\infty}$ converges to A in probability, denoted as $A_n \xrightarrow{(\mathrm{p})} A$ if for all $\delta > 0$ [21, Sec. 2.2],

$$\lim_{n\to\infty} \mathbb{P}[|A_n - A| > \delta] = 0. \tag{4}$$

The empirical distribution, or type, of a sequence $\boldsymbol{x} \in \mathcal{X}^n$, which will be denoted by $\hat{P}_{\boldsymbol{x}}$, is the vector of relative frequencies, $\hat{P}_{\boldsymbol{x}}(x)$, of each symbol $x \in \mathcal{X}$ in $\boldsymbol{x}$. The set of all possible empirical distributions of sequences of length n on alphabet $\mathcal{X}$ is denoted by $\mathcal{P}_n(\mathcal{X})$. The joint empirical distribution of a pair of sequences, denoted by $\hat{P}_{\boldsymbol{xy}}$, is similarly defined. The set of all possible joint empirical distributions of sequences of length n on alphabets $\mathcal{X}$ and $\mathcal{Y}$ is denoted by $\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$. The type class of Q_X, denoted by $\mathcal{T}(Q_X)$, is the set of all vectors $\boldsymbol{x} \in \mathcal{X}^n$ with $\hat{P}_{\boldsymbol{x}} = Q_X$. The joint type class of P_{XY}, denoted by $\mathcal{T}(P_{XY})$, is the set of pairs of sequences $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ with $\hat{P}_{\boldsymbol{xy}} = P_{XY}$. In addition, we also define $\mathcal{Q}(Q_X) \triangleq \{P_{XX'} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}) : P_X = P_{X'} = Q_X\}$. Finally, $[M]$ denotes the set $\{1, 2, \cdots, M\}$, and $[M]^2_* \triangleq \{(m,m') \in [M]^2 : m \neq m'\}$ for any M.

C. Structure of the Paper

In Section II, we introduce error probability and error exponents. In Section III-A, we introduce the generation of RGV random codebook ensembles. We also mention about properties of RGV codes and type-numerators in this section. We derive the typical error exponent for the RGV in Section IV. Finally, we study concentration properties of this ensemble in Section V. Proofs of the main results can be found in the corresponding sections while the proofs of auxiliary results can be found in the Appendices.

II. Preliminaries

We assume that a code $\mathcal{C}_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \in \mathcal{X}^n, M = e^{nR}$ is employed for transmission over a DMC channel with channel law $W(y|x)$ for $x \in \mathcal{X}, y \in \mathcal{Y}$. More specifically, when the transmitter wishes to convey a message $m \in \{1, 2, \cdots, M\}$, it sends codeword $\boldsymbol{x}_m = (x_{m,1}, \ldots, x_{m,n}) \in \mathcal{X}^n$ over the channel. The channel produces an output vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$, according to

$$W(\boldsymbol{y}|\boldsymbol{x}_m) = \prod_{i=1}^{n} W(y_i|x_{m,i}). \tag{5}$$

At the decoder side, we assume that a GLD [20] is used to infer what the transmitted message was. The GLD [20] extends the likelihood decoder in [22] and [23], and is a stochastic decoder that randomly selects the message estimate $\hat{m}$ according to the posterior probability distribution given the channel output $\boldsymbol{y}$ as follows

$$\Pr(\hat{m} = m|\boldsymbol{y}) = \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{y}})\}}{\sum_{m=1}^{M} \exp\left\{ng(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})\right\}}, \tag{6}$$

where $g(\cdot)$, henceforth referred to as the *decoding metric*, is an arbitrary continuous function of a joint distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$. For

$$g(P_{XY}) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P_{XY}(x, y) \log W(y|x), \tag{7}$$

we recover the ordinary likelihood decoder [23]. For

$$g(P_{XY}) = \beta \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P_{XY}(x, y) \log W(y|x), \tag{8}$$

$\beta \geq 0$ being a free parameter, we extend this to a parametric family of decoders, where β controls the skewness of the posterior [11]. In particular, $\beta \to \infty$ leads to the (deterministic) ML decoder, denoted by $g^{\text{ml}}(\cdot)$. Other interesting choices are associated with mismatched metrics,

$$g(P_{XY}) = \beta \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P_{XY}(x, y) \log W'(y|x), \tag{9}$$

W' being different from W, and

$$g^{\text{smi}}(P_{XY}) = \beta I_P(X;Y), \tag{10}$$

which is the stochastic version of the well-known universal maximum mutual information (MMI) decoder [24], which has been recently proven to be universal in a typical error exponent sense [17]. The MMI decoder is approached by letting $\beta \to \infty$ in (10).

The average probability of error, associated with a given code $\mathcal{C}_n$ and the GLD, is given by

$$P_{\text{e}}(\mathcal{C}_n) = \frac{1}{M} \sum_{m=1}^{M} \sum_{m'\neq m} \sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})\}}{\sum_{\tilde{m}=1}^{M} \exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}},\boldsymbol{y}})\}}. \tag{11}$$

The n-length error exponent of code $\mathcal{C}_n$ is defined as

$$E_n(\mathcal{C}_n) = -\frac{1}{n} \log P_{\text{e}}(\mathcal{C}_n). \tag{12}$$

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

For a given code ensemble, the RCE is defined as

$$E_{\text{rce}}(R, Q_X) \triangleq \liminf_{n\to\infty} -\frac{1}{n} \log \mathbb{E}[P_{\text{e}}(\mathcal{C}_n)] \tag{13}$$

For GLD, the RCE was derived by [23] (see also [20]) and is given by

$$\begin{aligned} &E_{\text{rce}}^{\text{cc}}(R, Q_X) \\ &= \min_{P_{XY}:P_X=Q_X} \min_{\tilde{P}_{XY}:\tilde{P}_X=Q_X,\tilde{P}_Y=P_Y} D(P_{XY}\|Q_X \times W) \\ &\quad + \Big[I_{\tilde{P}}(X, Y) + \big[\mathbb{E}_P[\log W(Y|X)] \\ &\quad - \mathbb{E}_{\tilde{P}}[\log W(Y|X)]\big]_+ - R\Big]_+ \end{aligned} \tag{14}$$

and was shown to coincide with the constant composition exponent for ML decoding.

For ML decoding, Csiszár and Körner [8] proved the existence of a constant composition code with exponent

$$E_{\text{ck}}^{\text{cc}}(R, Q_X) = \min_{P\in\mathcal{T}_I} D(P_{Y|X}\|W|P) + \big[I(X'; X, Y) - R\big]_+ \tag{15}$$

and

$$\begin{aligned} \mathcal{T}_{\text{ck}} = \Big\{&P_{XX'Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : P_X = P_{X'} = Q_X, \\ &\mathbb{E}_P[\log W(Y|X')] \geq \mathbb{E}_P[\log W(Y|X)], I_P(X; X') \leq R\Big\}. \end{aligned} \tag{16}$$

The Csiszár and Körner exponent, is known to be at least as large as the RCE and the expurgated exponent for constant composition codes derived by Csiszár, Körner and Marton [25] defined as

$$E_{\text{ckm}}^{\text{cc}}(R, Q_X) = \min_{I(X;X')\leq R} \mathbb{E}[d_{\text{B}}(X, X')] + I(X; X') - R, \tag{17}$$

where $d_{\text{B}}(\cdot, \cdot)$ is the Bhattacharyya distance defined as

$$d_{\text{B}}(x, x') = -\log \sum_{y\in\mathcal{Y}} \sqrt{W(y|x)W(y|x')}. \tag{18}$$

For GLD, Merhav provided an expression for the expurgated exponent for constant composition codes [20, Eq. (36)], given by

$$\begin{aligned} E_{\text{ex}}^{\text{cc}}(R, Q_X) = &\min_{P_{XX'}\in\mathcal{Q}(Q_X):I_P(X;X')\leq R} \Big\{\Gamma(P_{XX'}, R) \\ &+ I_P(X; X') - R\Big\}, \end{aligned} \tag{19}$$

where for $Q_X \in \mathcal{P}(\mathcal{X}), \Delta \in \mathbb{R}, d \in \Omega$, we define

$$\begin{aligned} \Gamma(P_{XX'}, R) \triangleq &\min_{P_{Y|XX'}} \Big\{D(P_{Y|X}\|W|Q_X) + I_P(X'; Y|X) \\ &+ [\max\{g(P_{XY}), \alpha(R, P_Y)\} - g(P_{X'Y})]_+\Big\}, \end{aligned} \tag{20}$$

and

$$\alpha(R, P_Y) \triangleq \max_{\substack{P_{X'|Y}: P_{X'}=Q_X, \\ I_P(X';Y)\le R}} \big(g(P_{X'Y}) - I_P(X';Y)\big) + R. \tag{21}$$

For a given code ensemble, the TRC defined as

$$E_{\rm trc}(R, Q_X) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\rm e}(\mathcal{C}_n)] \tag{22}$$

which is known to be strictly larger than the RCE for the same ensemble at low rates. In addition, Merhav also provided an expression for the TRC for the constant composition ensemble and GLD [11, Eq. (18)]

$$E^{\rm cc}_{\rm trc}(R, Q_X) = \min_{P_{XX'}\in\mathcal{Q}(Q_X): I_P(X;X')\le 2R} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\} \tag{23}$$

and showed that for GLD $E^{\rm cc}_{\rm trc}(R, Q_X) \le E^{\rm cc}_{\rm ex}(2R, Q_X) + R$; this inequality holds with equality for ML decoding.

In the next sections, we introduce RGV codebook ensemble and derive concentration properties of the error exponent (12) of sequences of RGV codes $\mathcal{C}_n$ in the asymptotic regime.

III. RGV Random Codebook Ensembles and Properties

A. RGV Random Codebook Ensemble

In this section, we describe basic RGV codebook construction as well as some of its properties. The RGV codebook was first introduced in [7], which extended code constructions that attain the Gilbert-Varshamov bound on the Hamming space [26], [27]. The RGV construction is a randomized constant composition counterpart of such codes for arbitrary DMCs and arbitrary distance functions.

Definition 1: Let Ω be the set of bounded, symmetric, and type-dependent functions $d(\cdot,\cdot): \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}$, i.e., bounded functions that satisfy $d(\boldsymbol{x}, \boldsymbol{x}') = d(\boldsymbol{x}', \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$, that depend on $(\boldsymbol{x}, \boldsymbol{x}')$ only through the joint distribution $\hat{P}_{\boldsymbol{x}\boldsymbol{x}'}$, and that are continuous on the probability simplex.

We refer to $d \in \Omega$ as a distance function, although it need not to be a distance in the topological space (e.g., it may be negative). Some examples of such distance function include Hamming distance, Bhattacharyya distance, and equivocation distance [7].

The RGV code $c_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \in \mathcal{X}^n$ with M codewords of length n is constructed such that any two distinct codewords $\boldsymbol{x}, \boldsymbol{x}' \in c_n$ satisfy $d(\boldsymbol{x}, \boldsymbol{x}') > \Delta$ for a given distance function $d(\cdot,\cdot) \in \Omega$ and $\Delta \in \mathbb{R}$. This guarantees that the minimum distance of the codebook exceeds the minimum distance Δ. The construction depends on the input distribution $Q_X \in \mathcal{P}_n(\mathcal{X})$ and is described by the following steps:

1) The first codeword, $\boldsymbol{x}_1$, is drawn equiprobably from $\mathcal{T}(Q_X)$;
2) The second codeword, $\boldsymbol{x}_2$, is drawn equiprobably from

$$\mathcal{T}(Q_X, \boldsymbol{x}_1) \triangleq \big\{\bar{\boldsymbol{x}} \in \mathcal{T}(Q_X) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_1) > \Delta\big\} \tag{24}$$
$$= \mathcal{T}(Q_X) \setminus \big\{\bar{\boldsymbol{x}} \in \mathcal{T}(Q_X) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_1) \le \Delta\big\}, \tag{25}$$

i.e., the set of sequences with composition Q_X whose distance to $\boldsymbol{x}_1$ exceeds Δ;
3) Continuing recursively, the i-th codeword $\boldsymbol{x}_i$ is drawn equiprobably from

$$\mathcal{T}(Q_X, \boldsymbol{x}_1^{i-1}) \triangleq \big\{\bar{\boldsymbol{x}} \in \mathcal{T}(Q_X) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_j) > \Delta, j = 1, 2, \ldots, i-1\big\} \tag{26}$$
$$= \mathcal{T}(Q_X, \boldsymbol{x}_1^{i-2}) \setminus \big\{\bar{\boldsymbol{x}} \in \mathcal{T}(Q_X, \boldsymbol{x}_1^{i-2}) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_{i-1}) \le \Delta\big\} \tag{27}$$

where for $j < k$, $\boldsymbol{x}_j^k = (\boldsymbol{x}_j, \ldots, \boldsymbol{x}_k)$ is a shorthand notation to denote previously drawn codewords.

This recursive procedure does not necessarily guarantee that $M = e^{nR}$ codewords have been obtained. As [7, Theorem 1], in order to ensure that the above procedure generates the desired number of codewords, it suffices to choose R such that, for some $\delta > 0$,

$$R \le \min_{P_{XX'}\in\mathcal{Q}(Q_X): d(P_{XX'})\le\Delta} I(X;X') - 2\delta. \tag{28}$$

For a given RGV code with rate R, type Q_X, distance function d, and minimum distance Δ, we define the RCE associated with decoding metric g as

$$E^{\rm rgv}_{\rm rce}(R, Q_X, g, d, \Delta) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log\mathbb{E}[P_{\rm e}(\mathcal{C}_n)] \tag{29}$$

and the TRC error exponent associated with decoding metric g as

$$E^{\rm rgv}_{\rm trc}(R, Q_X, g, d, \Delta) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\rm e}(\mathcal{C}_n)], \tag{30}$$

where the expectation is with respect to the randomness of the code $\mathcal{C}_n$.

The main result of [7] is that for ML decoding, and suitably optimized distance function and minimum distance, the RCE of the constant composition RGV ensemble is equal to the Csiszár and Körner exponent (15). In this paper, we study the TRC of the RGV ensemble with GLD. One of the main results of the paper is to provide a generic expression for $E^{\rm rgv}_{\rm trc}(R, Q_X, g, d, \Delta)$ as a function of the RGV code parameters. In addition, we show that

$$E^{\rm rgv}_{\rm trc}(R, Q_X, g, d, \Delta) = E^{\rm cc}_{\rm ex}(R, Q_X). \tag{31}$$

for a suitable choice of the RGV ensemble parameters. While $E^{\rm rgv}_{\rm rce}(R, Q_X, g, d, \Delta)$ potentially includes the asymptotic performance of relatively poor codes in the ensemble, $E^{\rm rgv}_{\rm trc}(R, Q_X, g, d, \Delta)$ provides the expected exponent. Hence, $E^{\rm rgv}_{\rm trc}(R, Q_X, g, d, \Delta)$ is the relevant exponent of interest. In addition, we provide bounds on the concentration rates of the lower and upper tails of the error exponent of RGV codes. We show that the lower tail decays exponentially while the upper tail decays double-exponentially.

B. Properties of RGV Codebooks

In this subsection, we introduce several technical results characterizing the key properties of the generalized RGV construction. We begin by restating some known properties from [7]; we will then introduce a number of other properties that will be helpful in the derivation of our main results.

Lemma 1 [7, Lemma 1]: Under condition (28), for some $\delta > 0$ and $\boldsymbol{x}_1^{i-1}$ occurring with non-zero probability (or $d(\boldsymbol{x}_k, \boldsymbol{x}_l) > \Delta, \forall k, l \in [i-1], k \neq l$), we have that

$$(1-e^{-n\delta})|\mathcal{T}(Q_X)| \le |\mathcal{T}(Q_X, \boldsymbol{x}_1^{i-1})| \le |\mathcal{T}(Q_X)|, \ \forall i \in [M]. \tag{32}$$

Lemma 2 [7, Lemma 2]: Under the condition (28), for any $k, m \in [M], k \neq m$ and $\boldsymbol{x}_k, \boldsymbol{x}_m \in \mathcal{T}(Q_X)$ such that $d(\boldsymbol{x}_k, \boldsymbol{x}_m) > \Delta$, then we have

$$\begin{aligned}\frac{1-4\delta_n^2}{|\mathcal{T}(Q_X)|^2} e^{-2\delta_n} &\le \mathbb{P}[\boldsymbol{X}_k = \boldsymbol{x}_k, \boldsymbol{X}_m = \boldsymbol{x}_m] \\ &\le \frac{1}{(1-e^{-n\delta})^2 |\mathcal{T}(Q_X)|^2}\end{aligned} \tag{33}$$

while $\mathbb{P}[\boldsymbol{X}_k = \boldsymbol{x}_k, \boldsymbol{X}_m = \boldsymbol{x}_m] = 0$ whenever $d(\boldsymbol{x}_k, \boldsymbol{x}_m) \le \Delta$, where,

$$\delta_n \triangleq \frac{e^{-n\delta}}{1-e^{-n\delta}}. \tag{34}$$

Lemma 3 [7, Lemma 4]: For any message index m, the marginal distribution of codeword $\boldsymbol{X}_m$ is $\mathbb{P}(\boldsymbol{x}_m) = \frac{1}{|\mathcal{T}(Q_X)|}$ for $\boldsymbol{x}_m \in \mathcal{T}(Q_X)$.

In order to derive the TRC and convergence properties of the RGV code ensemble, we need to derive new properties of this random codebook. Some properties of the pairwise independent fixed-composition code ensemble [11], [15] are proven to hold for the RGV codebook under some extra conditions by other proof techniques. First, the following lemma can be easily proved using the same arguments as [7].

Lemma 4: Consider the generalized RGV construction with the rate R satisfying (28). Then, for any $\mathcal{A} \subset [M]$ and any rate R satisfying (28) for some $\delta > 0$, under the condition that $\min_{k,l \in \mathcal{A}: k \neq l} d(\boldsymbol{x}_k, \boldsymbol{x}_l) > \Delta$, it holds that

$$\mathbb{P}\bigg[\bigcap_{k \in \mathcal{A}} \{\boldsymbol{X}_k = \boldsymbol{x}_k\}\bigg] \le \frac{1}{(1-e^{-n\delta})^{|\mathcal{A}|} |\mathcal{T}(Q_X)|^{|\mathcal{A}|}}. \tag{35}$$

In addition, if $\min_{k,l \in \mathcal{A}: k \neq l} d(\boldsymbol{x}_k, \boldsymbol{x}_l) \le \Delta$, it holds that

$$\mathbb{P}\bigg[\bigcap_{k \in \mathcal{A}} \{\boldsymbol{X}_k = \boldsymbol{x}_k\}\bigg] = 0. \tag{36}$$

Furthermore, if $\min_{k,l \in [M']: k \neq l} d(\boldsymbol{x}_k, \boldsymbol{x}_l) > \Delta$ for any $M' \le M$, it holds that

$$\mathbb{P}\bigg[\bigcap_{m \in [M']} \{\boldsymbol{X}_m = \boldsymbol{x}_m\}\bigg] \ge \frac{1}{|\mathcal{T}(Q_X)|^{M'}}. \tag{37}$$

In general, (37) does not hold for any $\mathcal{A} \subset [M]$ as (35), but it holds for the class of subsets $\{[M']\}_{M' \le M}$. If $\mathcal{A} = [M']$, we obtain both upper and lower bound on $\mathbb{P}\big[\bigcap_{m \in [M']} \{\boldsymbol{X}_m = \boldsymbol{x}_m\}\big]$.

Compared with Lemma 2, (37) is tighter at $M = 2$ if $\{k, m\} = \{1, 2\}$. However, Lemma 2 is more general, i.e., it holds for any subset $\{k, m\} : (k, m) \in [M] \times [M], \ k \neq m\}$.

Proof: See Appendix A. ■

Denote by

$$\mathcal{I}(m, m') \triangleq \mathbb{1}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})\}. \tag{38}$$

Then, the following result, whose proof can be found in Appendix B, holds.

Lemma 5: Let $P_{XX'}$ be a joint-type in $\mathcal{Q}(Q_X)$ such that $d(P_{XX'}) > \Delta$. Define

$$L(P_{XX'}) \triangleq \frac{|\mathcal{T}(P_{XX'})|}{|\mathcal{T}(Q_X)|^2}. \tag{39}$$

Then, under the condition (28) and $d(P_{XX'}) > \Delta$, for any two pairs $(i, j), (k, l) \in [M]_*^2$ such that $(i, j) \neq (k, l)$, it holds that

$$\begin{aligned}(1-4\delta_n^2) e^{-2\delta_n} L(P_{XX'}) &\le \mathbb{E}[\mathcal{I}(i, j)] \\ &\le \frac{1}{(1-e^{-n\delta})^2} L(P_{XX'}),\end{aligned} \tag{40}$$

and

$$\mathbb{E}[\mathcal{I}(i, j)\mathcal{I}(k, l)] \le \frac{1}{(1-e^{-n\delta})^4} L^2(P_{XX'}). \tag{41}$$

This implies that

$$\mathbb{E}[\mathcal{I}(i, j)] \doteq \exp\{-nI_P(X; X')\}, \tag{42}$$

$$\mathbb{E}[\mathcal{I}(i, j)\mathcal{I}(k, l)] \mathrel{\dot{\le}} \exp\{-2nI_P(X; X')\}. \tag{43}$$

C. Useful Properties of Type Enumerators

In this section, we state some important properties of the type enumerator of RGV codebooks. For a given joint-type $P_{XX'} \in \mathcal{Q}(Q_X)$, the type enumerator $N(P_{XX'})$ is defined as the number of codeword pairs with joint type $P_{XX'}$, i.e.,

$$\begin{aligned}N(P_{XX'}) &\triangleq \sum_m \sum_{m' \neq m} \mathbb{1}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})\} \qquad (44)\\ &= \sum_{(m, m') \in [M]_*^2} \mathcal{I}(m, m'), \qquad (45)\end{aligned}$$

where $\mathcal{I}(m, m')$ is defined in (38).

Lemma 6: Fix arbitrary small positive numbers $\delta > 0$ and $\varepsilon > 0$. Let $P_{XX'} \in \mathcal{Q}(Q_X)$ be a joint distribution that satisfies $I_P(X; X') < 2R - \varepsilon$ and $d(P_{XX'}) > \Delta$. Define

$$\begin{aligned}\mathcal{E}(P_{XX'}) = \bigg\{\mathcal{C}_n : N(P_{XX'}) &< (1-4\delta_n^2) e^{-2\delta_n} \\ &\times \exp\{n[2R - I_P(X; X') - \varepsilon]\}\bigg\}.\end{aligned} \tag{46}$$

Then, for any rate R satisfying (28), it holds (as n sufficiently large) that

$$\begin{aligned}&\mathbb{P}\big[\mathcal{E}(P_{XX'})\big] \\ &\le \frac{1}{(1-e^{-n\varepsilon/2})^2}\bigg[\frac{e^{4\delta_n}}{\big(1-4\delta_n^2\big)^2\big(1-e^{-n\delta}\big)^2} e^{-n\varepsilon/2} \\ &\quad + \frac{e^{4\delta_n}}{(1-4\delta_n^2)^2(1-e^{-n\delta})^4} - 1\bigg] \to 0\end{aligned} \tag{47}$$

as $n \to \infty$ for any fixed $\delta > 0$.

Proof: See Appendix C. ■

Lemma 7: Let $\varepsilon > 0$ be given and assume that the condition (28) holds. Then, for any $P_{XX'} \in \mathcal{Q}(Q_X)$ such that $I_P(X;X') \leq 2R$ and $d(P_{XX'}) > \Delta$,

$$\mathbb{P}\big[N(P_{XX'}) \geq e^{n(2R - I_P(X;X')+\varepsilon)}\big] \overset{\circ}{\leq} \exp\big\{-e^{n(2R-I_P(X;X')+\varepsilon)}\big\} \tag{48}$$
$$\dot{\leq} e^{-n\infty}. \tag{49}$$

Proof: See Appendix D. ■

Lemma 8: Let $\varepsilon > 0$ be given. Then, for any $P_{XX'} \in \mathcal{Q}(Q_X)$ such that $I_P(X;X') \geq 2R - \varepsilon$ and $d(P_{XX'}) > \Delta$ such that the condition (28) holds,

$$\mathbb{P}\big[N(P_{XX'}) \geq e^{n\varepsilon}\big] \overset{\circ}{\leq} \exp\big\{-e^{n\varepsilon}\big\} \tag{50}$$
$$\dot{\leq} e^{-n\infty}. \tag{51}$$

Proof: See Appendix E. ■

Lemma 9: For any $P_{XX'} \in \mathcal{Q}(Q_X)$ such that $I_P(X;X') \geq 2R$ and $d(P_{XX'}) > \Delta$ such that the condition (28) holds, we have

$$\mathbb{P}\big[N(P_{XX'}) \geq 1\big] \doteq \exp\{n(2R - I_P(X;X'))\}. \tag{52}$$

Proof: See Appendix F. ■

The following lemma is a key result for showing the exponentially-decay of the lower tail decay.

Lemma 10: Let $P_{XX'} \in \mathcal{Q}(Q_X)$ such that $d(P_{XX'}) > \Delta$. Then, under the condition (28), we have

$$\mathbb{P}\big[N(P_{XX'}) \geq e^{ns}\big] \doteq e^{-nE(R,P_{XX'},s)} \qquad \forall s \in \mathbb{R}, \tag{53}$$

where

$$E(R,P,s) = \begin{cases} [I_P(X;X') - 2R]_+, & [2R - I_P(X;X')]_+ > s \\ +\infty, & [2R - I_P(X;X')]_+ < s \end{cases}. \tag{54}$$

Proof: See a detailed proof in Appendix G. ■

The following lemma is a key enabling result to attain the double-exponential bound for the concentration properties of the random coding exponent in the RGV codebook. As opposed to the independent fixed-composition ensemble [15], a direct application of Suen's correlation inequality as [15, Proof of Lemma 2] does not give the double-exponential bound. More specifically, since all RGV codewords are correlated, the number of adjacent pairs of a fixed pair (m, m') is now e^{2nR} which causes the term in [15, Eq. (B.18)] to be equal to 1. For the independent fixed-composition code ensemble, this term is e^{nR}.

To overcome this difficulty, we develop another proof for this lemma which is not based on the Suen's correlation inequality. See Appendix H for a detailed proof.

Lemma 11: Let $\varepsilon > 0$ and $\mathcal{D} \subset \{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta\}$ be given. Then, under the condition

$$\min_{P_{XX'} \in \mathcal{D}} I_P(X;X') - 2\delta \leq R \leq \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \leq \Delta} I_P(X;X') - 2\delta, \tag{55}$$

or

$$R \leq \min\bigg\{ \min_{P_{XX'} \in \mathcal{D}} I_P(X;X') - \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \leq \Delta} I_P(X;X'), \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \leq \Delta} I_P(X;X')\bigg\} - 2\delta \tag{56}$$

for some $\delta > 0$, we have

$$\min_{P_{XX'} \in \mathcal{D}} \mathbb{P}\Big\{N(P_{XX'}) \leq e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\} \overset{\circ}{\leq} \exp\Big\{-\min\big(e^{n(R-2\delta)}, e^{n(2R - \min_{P_{XX'} \in D} I_P(X;X'))}\big)\Big\}. \tag{57}$$

Observe that for $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -(R + 2\delta)$, the condition (55) holds since

$$\min_{P_{XX'} \in \mathcal{D}} I_P(X;X') - 2\delta \leq \max_{P_{XX'}: d(P_{XX'}) > \Delta} I_P(X;X') - 2\delta \tag{58}$$
$$= \max_{P_{XX'}: I_P(X;X') < -\Delta} I_P(X;X') - 2\delta \tag{59}$$
$$< -(\Delta + 2\delta), \tag{60}$$

and

$$\min_{P_{XX'}: d(P_{XX'}) \leq \Delta} I_P(X;X') - 2\delta \tag{61}$$
$$= \min_{P_{XX'}: I_P(X;X') \geq -\Delta} I_P(X;X') - 2\delta \tag{62}$$
$$= -(\Delta + 2\delta). \tag{63}$$

Hence, the double-exponential expression in (57) holds for this special distance d and Δ. The condition (56) also holds for many other classes of distances d and different values of Δ.

Finally, we state the following key lemma, whose proof can be found in Appendix I.

Lemma 12: Recall the definition of $\Gamma(P_{XX'}, R)$ in (20). We define the expurgated error exponent for RGV ensemble as following:

$$E_{\text{ex}}^{\text{rgv}}(R, Q_X, g, d, \Delta) \triangleq \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) > \Delta, I_P(X;X') \leq R} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\}. \tag{64}$$

Let

$$\mathcal{A}_1 = \Big\{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, I_P(X;X') > 2R\Big\}, \tag{65}$$

$$\mathcal{A}_2 = \Big\{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, I_P(X;X') \leq 2R, \Gamma(P_{XX'}, R - \varepsilon) + I_P(X;X') - R \leq E_0 + \varepsilon\Big\}, \tag{66}$$

and define

$$\mathcal{F}_0 \triangleq \bigcap_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \big\{N(P_{XX'}) = 0\big\}. \tag{67}$$

Under the conditions that $R < E_{\text{ex}}^{\text{rgv}}(R, Q_X, g, d, \Delta)$ and

$$\min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \leq \Delta} I_P(X;X')$$

$$\geq \max_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta} I_P(X;X'), \tag{68}$$

$$R \leq \min_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})\leq\Delta} I_P(X;X') - 2\delta \tag{69}$$

for some $\delta > 0$, it holds that

$$\mathbb{P}(\mathcal{F}_0) \stackrel{\circ}{\geq} \exp\left\{-e^{n\max_{P_{XX'}\in\mathcal{A}_2}(2R-I_P(X;X')\delta)}\right\}. \tag{70}$$

Similarly to the preceeding discussion, setting $d(P_{XX'}) \triangleq -I_P(X;X')$, we obtain that

$$\min_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})\leq\Delta} I_P(X;X') \geq -\Delta, \tag{71}$$

$$\max_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta} I_P(X;X') < -\Delta, \tag{72}$$

so (68) holds. For (69) being hold, it is required that $R \leq -(\Delta + 2\delta)$.

In connection to Lemma 11, the proof of the related result in [15, Prep. 6] cannot be applied here since it uses the Suen's correlation inequality, i.e. [15, Fact 3]. Since all codewords in RGV ensemble are dependent, the number of adjacent nodes in the corresponding adjacency graph is too big which makes this type of arguments invalid. To overcome this difficulty, in Appendix I, we develop a new technique. However, the double-exponential constant in (70) is smaller than the one in [15, Prep. 6] for the fixed-composition code ensemble.

IV. Typical Random Coding Exponent of Gilbert-Varshamov Codes

In this section, we show an expression for the TRC of the RGV code ensemble. The expression, when optimized over the distance function $d(\cdot,\cdot)$ and minimum distance Δ, recovers Merhav's expurgated exponent for the GLD proposed in [20]. The main result, proven in Section IV-A, is stated in the following.

Theorem 1: Let $Q_X \in \mathcal{P}(\mathcal{X}), \Delta \in \mathbb{R}, d \in \Omega$. Recall the definitions of $\Gamma(P_{XX'}, R)$ and $\alpha(R, P_Y)$ in (20) and (21). Then, for any R satisfying the condition in (28), the typical random coding exponent of the RGV code ensemble with the GLD is given by

$$E_{\text{trc}}^{\text{rgv}}(R, Q_X, g, d, \Delta) = \min_{\substack{P_{X'|X}:P_{X'}=P_X=Q_X,\\ I_P(X;X')\leq 2R, d(P_{XX'})>\Delta}} \left\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\right\}. \tag{73}$$

Before proceeding with the proof of the result, some discussion is in order. Observe that if we remove the constraint $d(P_{XX'}) > \Delta$ (i.e., no constraint on the distance between each codeword pair), the expression of the TRC for the RGV ensemble code in (73) becomes the TRC of the constant composition code ensemble with composition Q_X under GLD decoding in [11, Eq. (18)]. In addition, as shown below, when the distance function $d(\cdot,\cdot)$ is optimized, and Δ is chosen appropriately, the TRC expression (73) recovers Merhav's expurgated $E_{\text{ex}}^{\text{cc}}(R, Q_X)$ defined in (19), which is at least as high as the maximum of the expurgated exponent and the random coding exponent.

The following results are similar to ones in [7, Section IV].

Corollary 1: Let $\varepsilon > 0$ be given, and let R, P, and $d \in \Omega$ be given. The TRC of the generalized RGV construction with sufficiently small δ, $d(P_{XX'}) = -I_P(X;X'), \Delta = -(R + 2\delta)$, sufficiently large n, and GLD rule is such that $E_{\text{trc}}^{\text{rgv}}(R, Q_X, g, d, \Delta) = E_{\text{ex}}^{\text{cc}}(R, Q_X)$, defined in (19).

Proof: First, it is easy to see that the choices $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -(R + 2\delta)$ are valid for all R in the sense of satisfying the rate condition in (28) (see proof of [7, Cor. 2]). Now, under the same choices, we have

$$E_{\text{trc}}^{\text{rgv}}(R, Q_X, g, d, \Delta)\Big|_{d(P_{XX'})=-I_P(X;X'),\Delta=-(R+2\delta)} \tag{74}$$

$$= \min_{\substack{P_{X'|X}:P_{X'}=Q_X,\\ I_P(X;X')\leq 2R, I_P(X;X')\leq R+2\delta}} \left\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\right\} \tag{75}$$

$$= \min_{\substack{P_{X'|X}:P_{X'}=Q_X,\\ I_P(X;X')\leq R+2\delta}} \left\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\right\}. \tag{76}$$

The result follows by taking $\delta \to 0$ and using the continuity of $E_{\text{trc}}^{\text{rgv}}(R, Q_X, g, d, \Delta)$ in R. ∎

Corollary 2: The TRC of the generalized RGV construction with sufficiently small δ, $d(P_{XX'}) = -I_P(X;X'), \Delta = -(R+2\delta)$, sufficiently large n particularized for ML decoding is such that

$$E_{\text{trc}}^{\text{rgv}}(R, Q_X, g^{\text{ml}}, d, \Delta)\Big|_{d(P_{XX'})=-I_P(X;X'),\Delta=-(R+2\delta)} \geq \max\left\{E_{\text{rce}}^{\text{cc}}(R, Q_X), E_{\text{ckm}}^{\text{cc}}(R, Q_X)\right\} \tag{77}$$

where

$$E_{\text{rce}}^{\text{cc}}(R, Q_X) = \min_{P_{Y|X}} D(P_{Y|X}\|W|Q_X) + [I(X;Y) - R]_+, \tag{78}$$

is the RCE for ML decoding and $E_{\text{ckm}}^{\text{cc}}(R, Q_X)$ is the Csiszár-Körner-Marton expurgated exponent defined in (17).

Proof: We lower bound $E_{\text{trc}}^{\text{rgv}}(R, Q_X, g, d, \Delta)$ for ML decoding by the typical error exponent for a sub-optimal GLD based on $g^{\text{smi}}(P) = I_P(X;Y)$, which is the stochastic mutual information decoder defined in (10). In this case, it can be ready verified that $\alpha(R, P_Y) = R$, which yields

$$\Gamma^{\text{smi}}(P_{XX'}, R) = \min_{P_{Y|XX'}} D(P_{Y|X}\|W\|Q_X) + I_P(X';Y|X) + [\max\{I_P(X;Y), R\} - I_P(X';Y)]_+. \tag{79}$$

Hence, we have

$$\begin{aligned} &E_{\text{trc}}^{\text{rgv}}(R, Q_X, g^{\text{ml}}, d, \Delta)\Big|_{d(P_{XX'})=-I_P(X;X'),\Delta=-R} \\ &\geq E_{\text{trc}}^{\text{rgv}}(R, Q_X, g^{\text{smi}}, d, \Delta)\Big|_{d(P_{XX'})=-I_P(X;X'),\Delta=-R} \quad (80)\\ &= \min_{P_{X'Y|X}:I_P(X;X')\leq R, P_{X'}=P_X=Q_X} D(P_{Y|X}\|W|Q_X) \\ &\quad + I_P(X';Y|X) + I_P(X;X') \\ &\qquad + [\max\{I_P(X;Y), R\} - I_P(X';Y)]_+ - R \quad (81)\\ &= \min_{P_{X'Y|X}:I_P(X;X')\leq R, P_{X'}=P_X=Q_X} D(P_{Y|X}\|W|Q_X) \\ &\quad + I_P(X';X|Y) + I_P(X';Y) \end{aligned}$$

$$+ [\max\{I_P(X;Y), R\} - I_P(X';Y)]_+ - R \tag{82}$$

$$= \min_{P_{X'Y|X}: I_P(X;X') \le R, P_{X'}=P_X=Q_X} D(P_{Y|X} \| W | Q_X) + I_P(X';X|Y) + [\max\{I_P(X;Y), I_P(X';Y)\} - R]_+ \tag{83}$$

$$\ge \min_{P_{X'Y|X}: I_P(X;X') \le R, P_{X'}=P_X=Q_X} D(P_{Y|X} \| W | Q_X) + I_P(X';X|Y) + [I_P(X;Y) - R]_+ \tag{84}$$

$$= E_{\rm rce}^{\rm cc}(R, Q_X), \tag{85}$$

where (80) follows from (76), and (81) follows from Theorem 1 and (79).

Similarly, by using the same arguments as [11, p.5], for ML decoding, we have

$$E_{\rm trc}^{\rm rgv}(R, Q_X, g, d, \Delta) = \inf_{P_{X'Y|X \in \mathcal{S}(R,Q_X)}} D(P_{Y|X} \| W | Q_X) + I_P(X';X,Y) - R \tag{86}$$

where

$$\mathcal{S}(R, Q_X) = \big\{P_{X'Y|X} : I_P(X;X') \le R, P_{X'} = P_X = Q_X, \mathbb{E}_P[\log W(Y|X')] \ge \max\{\mathbb{E}_P[\log W(Y|X)], a(R, P_Y)\}\big\} \tag{87}$$

and

$$a(R, P_Y) = \sup_{P_{X'|Y}: I_P(X';Y) \le R, P_X = P_{X'} = Q_X} \mathbb{E}_P[\log W(Y|X)]. \tag{88}$$

Then, we have

$$E_{\rm trc}^{\rm rgv}(R, Q_X, g^{\rm ml}, d, \Delta)\Big|_{d(P_{XX'}) = -I_P(X;X'), \Delta = -R}$$

$$\ge \inf_{P_{X'Y|X \in \mathcal{T}_{\rm ck}}} D(P_{Y|X} \| W | Q_X) + I_P(X';X,Y) - R \tag{89}$$

$$= E_{\rm ckm}^{\rm cc}(R, Q_X), \tag{90}$$

where $\mathcal{T}_{\rm ck}$ defined in (16). Here, (89) follows from (86), and (90) follows from [8, Lemma 4]. ■

The following proposition reveals that the above choice of (d, Δ) is a choice that maximizes the TRC given in Theorem 1.

Lemma 13: Under the setup of Theorem 1 with

$$R \le \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \le \Delta} I_P(X;X') - 2\delta \tag{91}$$

for some $\delta > 0$, we have

$$E_{\rm trc}^{\rm rgv}(R, Q_X, g, d, \Delta) \le E_{\rm trc}^{\rm rgv}(R, Q_X, g, d, \Delta)\Big|_{d(P_{XX'}) = -I_P(X;X'), \Delta = -(R+2\delta)}. \tag{92}$$

Proof: From (91), for all joint type $P_{XX'} \in \mathcal{Q}(Q_X)$ such that $d(P_{XX'}) \le \Delta$, we have $R + 2\delta \le I_P(X;X')$. Hence, if $R + 2\delta > I_P(X;X')$, it holds that $d(P_{XX'}) > \Delta$. This means that

$$\Big\{P_{XX'} \in \mathcal{Q}(Q_X) : I_P(X;X') < R + 2\delta\Big\}$$

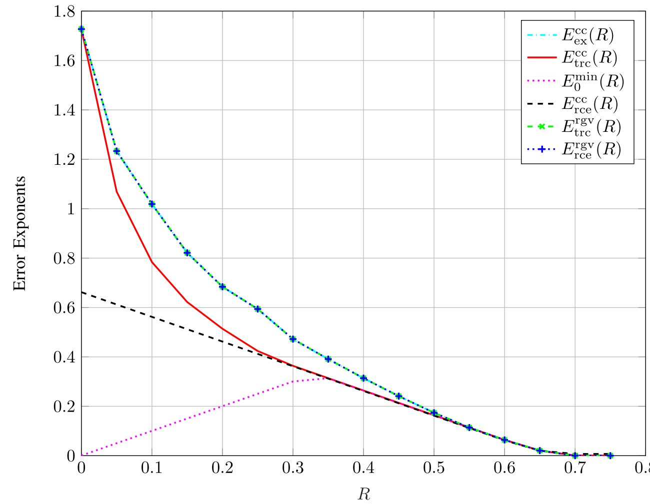

Fig. 1. Error exponents for the Z-channels with crossover probability 0.001 and ML decoding.

$$\subset \Big\{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta\Big\}. \tag{93}$$

It follows from (93) that for δ sufficiently small,

$$E_{\rm trc}^{\rm rgv}(R, Q_X, g, d, \Delta)$$

$$= \min_{\substack{P_{X'|X}: P_{X'} = P_X = Q_X, \\ I_P(X;X') \le 2R, d(P_{XX'}) > \Delta}} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\} \tag{94}$$

$$\le \min_{\substack{P_{X'|X}: P_{X'} = P_X = Q_X, \\ I_P(X;X') \le 2R, I_P(X;X') < R+2\delta}} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\} \tag{95}$$

$$= \min_{\substack{P_{X'|X}: P_{X'} = P_X = Q_X, \\ I_P(X;X') < R+2\delta}} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\} \tag{96}$$

$$= E_{\rm trc}^{\rm rgv}(R, Q_X, g, d, \Delta)\Big|_{d(P_{XX'}) = -I_P(X;X'), \Delta = -(R+2\delta)}, \tag{97}$$

where (97) follows from the continuity of $E_{\rm trc}^{\rm rgv}(R, Q_X, g, d, \Delta)$ in R and (76). ■

As in [7], the choice $d(P_{XX'}) = -I_P(X;X')$ is universally optimal in maximizing the TRC in Theorem 1 (subject to (28)), in the sense that it does not depend on the channel or input distribution.

In Fig. 1, we plot various error exponents for the Z-channel with crossover probability 0.001 and let $Q_X(0) = Q_X(1) = 1/2$. This example was considered in [15] and [20]. Specifically, for reference we plot the random coding exponent $E_{\rm rce}^{\rm cc}(R)$, the expurgated exponent $E_{\rm ex}^{\rm cc}(R)$, and the TRC $E_{\rm trc}^{\rm cc}(R)$ for constant composition codes. For the RGV ensemble exponents, we choose $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -R$ so as to achieve the largest possible exponents. We plot the corresponding random coding exponent $E_{\rm rce}^{\rm rgv}(R)$ and its corresponding TRC $E_{\rm trc}^{\rm rgv}(R)$ and illustrate that they both coincide with Merhav's expurgated exponent $E_{\rm ex}^{\rm cc}(R)$.

A. Proof of Theorem 1

The proofs for both upper and lower bounds follow similar lines to those in [11]. The main difference is the dependence among codeword induced by the RGV ensemble. In order to analyze this dependence, we developed new concentration inequalities and applied generalized versions of Hoeffding's inequality.

1) Lower Bound on TRC: First, we prove the following result.

Lemma 14: Recall the definition of $\alpha(R, P_Y)$ in (21). Fix an $\varepsilon > 0$. For any $m \in [M]$, let

$$Z_m(\boldsymbol{y}) \triangleq \sum_{\tilde{m}\neq m} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}},\boldsymbol{y}})}. \tag{98}$$

and

$$\mathcal{A}_m \triangleq \{Z_m(\boldsymbol{y}) \le \exp\{n\alpha(R-\varepsilon, \hat{P}_{\boldsymbol{y}})\}\}. \tag{99}$$

Then, under the condition (28), it holds that

$$\mathbb{P}\big[\mathcal{A}_m\big] \;\dot{\le}\; \exp\bigg\{-e^{n\varepsilon}\bigg[1-\frac{e^{-n(\varepsilon+\delta)}}{1-e^{-n\delta}} - e^{-n\varepsilon}(1+n\varepsilon)\bigg]\bigg\} \tag{100}$$

for all $m \in [M]$.

Proof: See Appendix K. ■

Proposition 1: Under the same assumptions as Theorem 1, the RGV code ensemble satisfies

$$\begin{aligned}&E_{\mathrm{trc}}^{\mathrm{rgv}}(R, Q_X, g, d, \Delta)\\&\ge \min_{\substack{P_{XX'}\in\mathcal{Q}(Q_X),\\ I_P(X;X')\le 2R, d(P_{XX'})>\Delta}} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\}.\end{aligned} \tag{101}$$

Proof: Using the GLD, the error probability is

$$\begin{aligned}P_{\mathrm{e}}(\mathcal{C}_n) &= \frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m)\\&\times \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{Y}})\} + \sum_{\tilde{m}\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}},\boldsymbol{y}})\}}.\end{aligned} \tag{102}$$

From (102), we obtain

$$\begin{aligned}&\mathbb{E}\big[P_{\mathrm{e}}(\mathcal{C}_n)\big]\\&\le \mathbb{E}\bigg[\frac{1}{M}\sum_{m=1}^{M}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{X}_m)\sum_{m'\neq m}\min\bigg\{1,\\&\qquad \frac{e^{ng(\hat{P}_{\boldsymbol{X}_{m'},\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{y}})} + \sum_{\tilde{m}\neq m} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}},\boldsymbol{y}})}}\bigg\}\bigg] \qquad (103)\\&= \mathbb{E}\bigg[\frac{1}{M}\sum_{m=1}^{M}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{X}_m)\sum_{\substack{m'\neq m\\ d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta}}\min\bigg\{1,\\&\qquad \frac{e^{ng(\hat{P}_{\boldsymbol{X}_{m'},\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{y}})} + \sum_{\tilde{m}\neq m} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}},\boldsymbol{y}})}}\bigg\}\bigg], \qquad (104)\end{aligned}$$

where (104) follows from the fact that $\min_{m'\neq m} d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta$ for any code $\mathcal{C}_n = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M)$ in the RGV codebook ensemble.

Now, we use similar arguments as [11] with some changes to cooperate the condition $d(x_m, x_{\tilde{m}}) > \Delta$ in the sum in (104). From (104) and Lemma 14, for any $\varepsilon > 0$, we obtain

$$\begin{aligned}&\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\\&\le \mathbb{E}\bigg[\frac{1}{M}\sum_{m=1}^{M}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{X}_m)\sum_{\substack{m'\neq m\\ d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta}}\min\bigg\{1,\\&\qquad \frac{e^{ng(\hat{P}_{\boldsymbol{X}_{m'},\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{y}})} + e^{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})}}\bigg\}\bigg].\end{aligned} \tag{105}$$

From the method of types [28] we have that

$$W(\boldsymbol{y}|\boldsymbol{x}_{\tilde{m}}) = e^{-n\left[H(\hat{P}_{\boldsymbol{x}_{\tilde{m}},y}) - H(Q_X) + D\left(\hat{P}_{\boldsymbol{x}_{\tilde{m}},y}\|Q_X\times W\right)\right]}. \tag{106}$$

Thus, it follows from (106) that

$$\begin{aligned}&\sum_{m=1}^{M}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\sum_{\substack{m'\neq m\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta}}\min\bigg\{1,\\&\qquad \frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{y}})} + e^{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})}}\bigg\} \qquad (107)\\&\doteq \sum_{m=1}^{M}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\sum_{\substack{m'\neq m\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta}}\exp\Big\{-n\big[\max\{g(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{y}}),\\&\qquad \alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\} - g(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})\big]_+\Big\} \qquad (108)\\&= \sum_{m=1}^{M}\sum_{\boldsymbol{y}}\sum_{\substack{m'\neq m\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta}}\exp\big\{\big(-n\big[H(\hat{P}_{\boldsymbol{x}_m,y}) - H(Q_X)\\&\quad + D\big(\hat{P}_{\boldsymbol{x}_m,y}\|Q_X\times W\big)\big]\big)\big\}\exp\Big\{-n\big[\max\{g(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{y}}),\\&\qquad \alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\} - g(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})\big]_+\Big\} \qquad (109)\\&\doteq \sum_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta} N(P_{XX'})\\&\quad\times\sum_{P_{Y|XX'}}\exp\{nH_P(Y|XX')\}\exp\Big\{\big(-n\big[H(P_{XY})\\&\quad - H(Q_X) + D\big(P_{XY}\|Q_X\times W\big)\big]\big)\Big\}\\&\quad\times\exp\Big\{-n[\max\{g(P_{XY}),\\&\qquad \alpha(R-\varepsilon,P_Y)\} - g(P_{X'Y})]_+\Big\} \qquad (110)\\&\doteq \sum_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta} N(P_{XX'})\\&\quad\times\exp\Big\{-n\min_{P_{Y|XX'}}\Big(-H_P(Y|XX') + H(P_{XY})\\&\quad - H(Q_X) + D\big(P_{XY}\|Q_X\times W\big)\\&\quad + \big[\max\{g(P_{XY}),\alpha(R-\varepsilon,P_Y)\} - g(P_{X'Y})\big]_+\Big)\Big\} \qquad (111)\\&\doteq \sum_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta} N(P_{XX'})\end{aligned}$$

$$\times \exp\Big\{ - n \min_{P_{Y|XX'}} \big(D(P_{Y|X}\|W|Q_X) + I_P(X';Y|X) + \big[\max\{g(P_{XY}), \alpha(R-\varepsilon, P_Y)\} - g(P_{X'Y})\big]_+\big)\Big\} \tag{112}$$

$$= \sum_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'})>\Delta} N(P_{XX'}) \times \exp\{-n\Gamma(P_{XX'}, R-\varepsilon)\}, \tag{113}$$

where (109) follows from (106), and (113) follows from (20). Here, the joint type enumerator $N(P_{XX'})$ has been defined in (45). From (105), (113), and (45), we obtain

$$\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{C}_n)] \leq \log\Big(\mathbb{E}\big[P_{\mathrm{e}}(\mathcal{C}_n)\big]\Big) \tag{114}$$

$$\leq \log\Bigg(\sum_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'})>\Delta} \mathbb{E}\big[N(P_{XX'})\big] \times \exp\big\{ - n\Gamma(P_{XX'}, R)\big\}\Bigg) - nR, \tag{115}$$

where (114) follows from the concavity of $\log x$ in $(0, \infty)$ and Jensen's inequality.

Now, for any $P_{XX'} \in \mathcal{Q}(Q_X)$ such that $d(P_{XX'}) > \Delta$, from Lemma 5,we obtain

$$\mathbb{E}\big[N(P_{XX'})\big] = \sum_{m=1}^{M} \sum_{m' \neq m} \mathbb{P}\big[(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\big] \tag{116}$$

$$\doteq e^{n(2R - I_P(X;X'))}. \tag{117}$$

Hence, from (115) and (117), we obtain

$$\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{C}_n)] \,\dot{\leq}\, \log\Bigg(\sum_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'})>\Delta} e^{n(2R - I_P(X;X'))} \times \exp\big\{ - n\Gamma(P_{XX'}, R)\big\}\Bigg) - nR. \tag{118}$$

From (118), we finally have

$$E_{\mathrm{trc}}^{\mathrm{rgv}}(R, Q_X, g, d, \Delta) \geq \min_{\substack{P_{XX'}: P_{X'} = P_X, \\ I_P(X;X') \leq 2R, d(P_{XX'})>\Delta}} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\}. \tag{119}$$

This concludes the proof of Proposition 1. ∎

2) Upper Bound on TRC:

Proposition 2: Under the same assumptions as Theorem 1, the RGV code ensemble satisfies

$$E_{\mathrm{trc}}^{\mathrm{rgv}}(R, Q_X, d, \Delta) \leq \min_{\substack{P_{XX'} \in \mathcal{Q}(Q_X): \\ I_P(X;X') \leq 2R, d(P_{XX'})>\Delta}} \Big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\Big\}. \tag{120}$$

Proof: The following proof follows similar lines to the proof in [11, Sect. 5.2]. However, the same proof cannot be used for the RGV ensemble. In addition to the difference in proofs of Lemmas 6 and (159), we also need to make additional changes in since the decay rate of $\mathbb{P}[\mathcal{E}(P_{XX'})]$ in Lemma 6 is not exponential as [11, Eq. (48)].

Given a joint-type $P_{XX'} \in \mathcal{Q}(Q_X)$ such that $I_P(X;X') < 2R - \varepsilon$ and $d(P_{XX'}) > \Delta$, let us define

$$Z_{mm'}(\boldsymbol{y}) = \sum_{\tilde{m} \neq m, m'} \exp\{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}})\}, \tag{121}$$

and

$$\mathcal{G}_n(P_{Y|XX'}) = \Big\{\mathcal{C}_n : \sum_m \sum_{m' \neq m} \mathcal{I}(m, m') \times \sum_{\boldsymbol{y} \in \mathcal{T}(P_{Y|XX'})} \mathbb{1}\{Z_{mm'}(\boldsymbol{y}) \leq \exp\{n[\alpha(R + 2\varepsilon, P_Y) + \varepsilon]\}\} \geq (1 - 4\delta_n^2)e^{-2\delta_n} \exp\{n[2R - I_P(X;X') - 3\varepsilon/2]\} \times |\mathcal{T}(P_{Y|XX'})|\Big\}, \tag{122}$$

where $\mathcal{I}(m, m')$ is defined in (38). Recall the definition of $\mathcal{E}(P_{XX'})$ in Eq. (46) Lemma 6. Then, similarly to [11, Sect. 5.2] we have

$$\mathbb{P}\big[\mathcal{G}_n^c(P_{Y|XX'}) \cap \mathcal{E}^c(P_{XX'})\big] \leq \mathbb{P}\Big[\sum_m \sum_{m' \neq m} \mathcal{I}(m, m') \sum_{\boldsymbol{y} \in \mathcal{T}(P_{Y|XX'})} \mathbb{1}\{Z_{mm'}(\boldsymbol{y}) \leq (1 - 4\delta_n^2)e^{-2\delta_n} \exp\{n[\alpha(R + 2\varepsilon, P_Y) + \varepsilon]\}\} \leq \exp\{n[2R - I_P(X;X') - 3\varepsilon/2]\}.|\mathcal{T}(P_{Y|XX'})|, N(P_{XX'}) \geq (1 - 4\delta_n^2)e^{-2\delta_n} \exp\{n[2R - I_P(X;X') - \varepsilon]\}\Big] \tag{123}$$

$$\leq \mathbb{P}\Big[\sum_m \sum_{m' \neq m} \mathcal{I}(m, m') \sum_{\boldsymbol{y} \in \mathcal{T}(P_{Y|XX'})} \mathbb{1}\{Z_{mm'}(\boldsymbol{y}) > (1 - 4\delta_n^2)e^{-2\delta_n} \exp\{n[\alpha(R+2\varepsilon, P_Y)+\varepsilon]\}\} \geq \big(\exp\{n[2R - I_P(X;X') - \varepsilon]\} - \exp\{n[2R - I_P(X;X') - 3\varepsilon/2]\}\big).|\mathcal{T}(P_{Y|XX'})|, N(P_{XX'}) \geq (1 - 4\delta_n^2)e^{-2\delta_n} \exp\{n[2R - I_P(X;X') - \varepsilon]\}\Big] \tag{124}$$

$$\leq \mathbb{P}\Big[\sum_m \sum_{m' \neq m} \mathcal{I}(m, m') \sum_{\boldsymbol{y} \in \mathcal{T}(P_{Y|XX'})} \mathbb{1}\{Z_{mm'}(\boldsymbol{y}) > (1 - 4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R + 2\varepsilon, P_{XY}) + \varepsilon]\}\} \geq \big(\exp\{n[2R - I_P(X;X') - \varepsilon]\} - \exp\{n[2R - I_P(X;X') - 3\varepsilon/2]\}\big).|\mathcal{T}(P_{Y|XX'})|\Big] \tag{125}$$

$$= \frac{\sum_m \sum_{m' \neq m} \sum_{\boldsymbol{y} \in \mathcal{T}(P_{Y|XX'})} \zeta(m, m', \boldsymbol{y})}{\exp\{n[2R - I_P(X;X') - \varepsilon]\}|\mathcal{T}(P_{Y|XX'})|}, \tag{126}$$

where (126) follows from Markov's inequality and

$$\zeta(m, m', \boldsymbol{y}) \triangleq \mathbb{P}\big[(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'}), Z_{mm'}(\boldsymbol{y}) > (1 - 4\delta_n^2)e^{-2\delta_n} \exp\{n[\alpha(R + 2\varepsilon, P_Y) + \varepsilon]\}\big] \tag{127}$$

$$= \sum_{\substack{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'}):\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta}} \mathbb{P}(\boldsymbol{x}_m,\boldsymbol{x}_{m'}) \times \mathbb{P}\big[Z_{mm'}(\boldsymbol{y})>(1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y)+\varepsilon]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big]. \tag{128}$$

Here, (128) follows from the fact that $\mathbb{P}(\boldsymbol{x}_m,\boldsymbol{x}_{m'})=0$ if $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})<\Delta$ by Lemma 2.

Now, given a fixed pair $(\boldsymbol{x}_m,\boldsymbol{x}_{m'})$ such that $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta$, define

$$P^*_{X'|Y} \triangleq \arg\max_{P_{X'|Y}} \mathbb{P}\big[N(P_{X'|Y}) > (n+1)^{-|\mathcal{X}||\mathcal{Y}|}(1-4\delta_n^2)e^{-2\delta_n}\exp\{n[\alpha(R+2\varepsilon,P_Y) + \varepsilon - g(P_{X'Y})]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big], \tag{129}$$

where

$$N(P_{X'|Y}) := \sum_{\tilde{m}\neq m,m'} \mathbb{1}\big\{(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{\tilde{X}Y})\big\}. \tag{130}$$

Then, we have

$$\mathbb{P}\big[Z_{mm'}(\boldsymbol{y})>(1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y)+\varepsilon]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big]$$

$$= \mathbb{P}\Bigg[\sum_{\tilde{m}\neq m,m'}\exp\{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}},\boldsymbol{y}})\}>(1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y)+\varepsilon]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\Bigg] \tag{131}$$

$$\leq \mathbb{P}\Bigg[\sum_{P_{X'|Y}} N(P_{X'|Y})\exp\{ng(P_{X'Y})\} > (1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y)+\varepsilon]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\Bigg] \tag{132}$$

$$= \mathbb{P}\Bigg[\sum_{P_{X'|Y}} N(P_{X'Y})\exp\{ng(P_{X'Y})\} > (1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y)+\varepsilon]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\Bigg] \tag{133}$$

$$\doteq \max_{P_{X'|Y}} \mathbb{P}\big[N(P_{X'|Y})\exp\{ng(P_{X'Y})\} > (n+1)^{-|\mathcal{X}||\mathcal{Y}|}(1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y)+\varepsilon]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big] \tag{134}$$

$$= \max_{P_{X'|Y}} \mathbb{P}\big[N(P_{X'|Y}) > (n+1)^{-|\mathcal{X}||\mathcal{Y}|}(1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y) + \varepsilon - g(P_{X'Y})]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big] \tag{135}$$

$$= \mathbb{P}\big[N(P^*_{X'|Y}) > (n+1)^{-|\mathcal{X}||\mathcal{Y}|}(1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y) + \varepsilon - g(P^*_{X'Y})]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big], \tag{136}$$

where (136) follows from (129).

Now, for all $\tilde{m}\in[M]$, observe that

$$\mathbb{P}\big[(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{X'Y})\big] = \sum_{\boldsymbol{x}_{\tilde{m}}\in\mathcal{T}(P_{X'|Y})}\mathbb{P}(\boldsymbol{x}_{\tilde{m}}) \tag{137}$$

$$= \frac{|\mathcal{T}(P_{X'|Y})|}{|\mathcal{T}(Q_X)|} \tag{138}$$

$$:= p, \tag{139}$$

where (138) follows from Lemma [7, Lemma 4]. It is easy to see that p does not depend on $\tilde{m}$.

Now, we consider two cases:

Case 1: $I_{P^*}(X';Y)\leq R+2\varepsilon$. Then, we have

$$\alpha(R+2\varepsilon,P_Y)+\varepsilon-g(P^*_{X'Y}) = \max_{\substack{P_{X'|Y}:P_{X'}=Q_X,\\ I_P(X';Y)\leq R+2\varepsilon}}\big(g(P_{X'Y})-I_P(X';Y)\big) + R+2\varepsilon-g(P^*_{X'Y}) \tag{140}$$

$$\geq g(P^*_{X'Y})-I_{P^*}(X';Y)+R+2\varepsilon-g(P^*_{X'Y}) \tag{141}$$

$$= R+2\varepsilon-I_{P^*}(X';Y). \tag{142}$$

On the other hand, if we let

$$\gamma \triangleq \frac{p}{1-e^{-n\delta}}, \tag{143}$$

we have

$$(M-2)\gamma \doteq \frac{e^{n(R-I_{P^*}(X';Y))}}{1-e^{-n\delta}} \tag{144}$$

It follows that

$$\mathbb{P}\big[N(P^*_{X'Y}) > (n+1)^{-|\mathcal{X}||\mathcal{Y}|}(1-4\delta_n^2)e^{-2\delta_n} \times \exp\{n[\alpha(R+2\varepsilon,P_Y) + \varepsilon - g(P^*_{X'Y})]\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big]$$
$$\dot{\leq} \mathbb{P}\big[N(P^*_{X'Y})>(M-2)\gamma e^{2n\varepsilon}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big] \tag{145}$$

where the last step follows from (144) and (142). Now, let $Z_{\tilde{m}}\triangleq\mathbb{1}\{(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{X'Y})\}$. Then, for all $\mathcal{A}\subset[M]\setminus\{m,m'\}$, under the condition (28), by Lemma 4, it holds that

$$\mathbb{E}\Bigg[\prod_{\tilde{m}\in\mathcal{A}}Z_{\tilde{m}}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\Bigg] = \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{|\mathcal{A}|}}\prod_{\tilde{m}\in\mathcal{A}}\mathbb{1}\{(\boldsymbol{x}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{X'Y})\} \times \mathbb{P}\Bigg[\bigcap_{\tilde{m}\in\mathcal{A}}\{\boldsymbol{X}_{\tilde{m}}=\boldsymbol{x}_{\tilde{m}}\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\Bigg]. \tag{146}$$

Now, observe that

$$\mathbb{P}\Bigg[\bigcap_{\tilde{m}\in\mathcal{A}}\{\boldsymbol{X}_{\tilde{m}}=\boldsymbol{x}_{\tilde{m}}\}\big|\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\Bigg] = \frac{\mathbb{P}\big(\bigcap_{\tilde{m}\in\mathcal{A}\cup\{m,m'\}}\{\boldsymbol{X}_{\tilde{m}}=\boldsymbol{x}_{\tilde{m}}\}\big)}{\mathbb{P}\big(\boldsymbol{X}_m=\boldsymbol{x}_m,\boldsymbol{X}_{m'}=\boldsymbol{x}_{m'}\big)} \tag{147}$$

$$\leq \frac{1}{(1-e^{-\delta n})^{|\mathcal{A}|+2}}\bigg(\frac{1}{|\mathcal{T}(Q_X)|^{|\mathcal{A}|+2}}\bigg)\frac{|\mathcal{T}(Q_X)|^2}{1-4\delta_n^2}e^{2\delta_n} \tag{148}$$

where (148) follows from Lemma 4 and Lemma 2 with noting that $d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta$.

Hence, it holds that

$$\begin{aligned}
&\mathbb{E}\bigg[\prod_{\tilde{m}\in\mathcal{A}} Z_{\tilde{m}} \big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\bigg] \\
&\le \left(\frac{e^{2\delta_n}}{1-4\delta_n^2}\right)\left(\frac{1}{(1-e^{-\delta n})^{|\mathcal{A}|+2}}\right) \\
&\quad \times \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{|\mathcal{A}|}} \prod_{\tilde{m}\in\mathcal{A}} \mathbb{P}[\boldsymbol{X}_{\tilde{m}} = \boldsymbol{x}_{\tilde{m}}] \\
&\quad \times \prod_{\tilde{m}\in\mathcal{A}} \mathbb{1}\{(\boldsymbol{x}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y})\} \qquad (149) \\
&= \left(\frac{e^{2\delta_n}}{1-4\delta_n^2}\right)\left(\frac{1}{(1-e^{-\delta n})^{|\mathcal{A}|+2}}\right) \prod_{\tilde{m}\in\mathcal{A}} \sum_{\boldsymbol{x}_{\tilde{m}}} \mathbb{P}[\boldsymbol{X}_{\tilde{m}} = \boldsymbol{x}_{\tilde{m}}] \\
&\quad \times \mathbb{1}\{(\boldsymbol{x}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y})\} \qquad (150) \\
&= \left(\frac{e^{2\delta_n}}{1-4\delta_n^2}\right)\left(\frac{1}{(1-e^{-\delta n})^{|\mathcal{A}|+2}}\right) \\
&\quad \times \prod_{\tilde{m}\in\mathcal{A}} \mathbb{P}\big[(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y})\big] \qquad (151) \\
&\doteq \left(\frac{p}{1-e^{-\delta n}}\right)^{|\mathcal{A}|}, \qquad (152)
\end{aligned}$$

where (149) follows from Lemma 4 (under the condition (28)). Hence, by applying Lemma 20, we have

$$\begin{aligned}
&\mathbb{P}\big[N(P^*_{X'Y}) > (M-2)\gamma e^{2n\varepsilon} \big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big] \\
&\quad \dot{\le} \exp\Big\{-e^{nR} D(e^{-na}\|e^{-nb})\Big\} \qquad (153)
\end{aligned}$$

where $D(p\|q)$ is the relative entropy between two Bernouilli distributions, with success probability p, q, respectively, and $a \triangleq I_{P^*}(X';Y) - 2\varepsilon + (1/n)\log(1-e^{-n\delta})$ and $b \triangleq I_{P^*}(X';Y) + (1/n)\log(1-e^{-n\delta})$. Since $b - a = 2\varepsilon$, by using the following fact [29, Sec. 6.3]:

$$D(a\|b) \ge a \log\frac{a}{b} + b - a, \qquad (154)$$

we have

$$\begin{aligned}
D\big(e^{-an}\|e^{-bn}\big) &\ge e^{-bn}\big[1 + e^{(b-a)n}((b-a)n - 1)\big] \qquad (155) \\
&\doteq e^{-nI_{P^*}(X';Y)} e^{2n\varepsilon} 2n\varepsilon. \qquad (156)
\end{aligned}$$

From (153) and (156), for any pair $(\boldsymbol{x}_m, \boldsymbol{x}_{m'})$ such that $d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta$, we obtain

$$\begin{aligned}
&\mathbb{P}\big[N(P^*_{X'Y}) > (M-2)\gamma e^{2n\varepsilon} \big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big] \\
&\dot{\le} \exp\Big\{-e^{n(R - I_{P^*}(X';Y))} e^{2n\varepsilon} 2n\varepsilon\Big\} \qquad (157) \\
&\le \exp\Big\{-e^{-2n\varepsilon} e^{2n\varepsilon} 2n\varepsilon\Big\} \qquad (158) \\
&= \exp\big\{-2n\varepsilon\big\}, \qquad (159)
\end{aligned}$$

where (158) follows from the condition $I_{P^*}(X';Y) \le R + 2\varepsilon$.

Case 2: $I_{P^*}(X';Y) > R + 2\varepsilon$. For this case, for any pair $(\boldsymbol{x}_m, \boldsymbol{x}_{m'})$ such that $d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta$, we have

$$\begin{aligned}
&\mathbb{P}\big[N(P^*_{X'Y}) > (n+1)^{-|\mathcal{X}||\mathcal{Y}|}(1-4\delta_n^2)e^{-2\delta_n} \\
&\quad \times \exp\{n[\alpha(R+2\varepsilon, P_Y) \\
&\quad + \varepsilon - g(P^*_{X'Y})]\}\big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big] \\
&\le \mathbb{P}\big[N(P^*_{X'Y}) \ge 1 \big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big] \qquad (160) \\
&\le \mathbb{E}[N(P^*_{X'Y}) \big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}] \qquad (161) \\
&= \sum_{\tilde{m}\neq m,m'} \mathbb{P}\big((\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y}) \big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big), \qquad (162)
\end{aligned}$$

where (160) follows from the fact that $N(P^*_{X'Y}) \in \mathbb{Z}_+$, and (161) follows from the Markov's inequality.

Now, by using (148) with $\mathcal{A} = \{\tilde{m}\}$, we have

$$\begin{aligned}
&\mathbb{P}\Big[\{\boldsymbol{X}_{\tilde{m}} = \boldsymbol{x}_{\tilde{m}}\} \big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\Big] \\
&\quad \le \frac{1}{(1-e^{-\delta n})^3}\left(\frac{1}{|\mathcal{T}(Q_X)|^3}\right)\frac{|\mathcal{T}(Q_X)|^2}{1-4\delta_n^2} e^{2\delta_n} \qquad (163) \\
&\quad \doteq \frac{1}{|\mathcal{T}(Q_X)|}. \qquad (164)
\end{aligned}$$

From (162) and (164), for any pair $(\boldsymbol{x}_m, \boldsymbol{x}_{m'})$ such that $d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta$, we obtain

$$\begin{aligned}
&\mathbb{P}\big[N(P^*_{X'Y}) > (n+1)^{-|\mathcal{X}||\mathcal{Y}|}(1-4\delta_n^2)e^{-2\delta_n} \\
&\quad \times \exp\{n[\alpha(R+2\varepsilon, P_Y) \\
&\quad + \varepsilon - g(P^*_{X'Y})]\}\big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big] \\
&\quad \dot{\le} (M-2)p \qquad (165) \\
&\quad \doteq e^{n(R - I_{P^*}(X';Y))} \qquad (166) \\
&\quad \le e^{-2n\varepsilon}, \qquad (167)
\end{aligned}$$

where (166) follows from (139), and (167) follows from condition $I_{P^*}(X';Y) > R + 2\varepsilon$.

From (159) and (167), for any pair $(\boldsymbol{x}_m, \boldsymbol{x}_{m'})$ such that $d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta$, we have

$$\begin{aligned}
&\mathbb{P}\big[N(P^*_{X'Y}) > (n+1)^{-|\mathcal{X}|}(1-4\delta_n^2)e^{-2\delta_n} \\
&\quad \times \exp\{n[\alpha(R+2\varepsilon, P_Y) \\
&\quad\quad + \varepsilon - g(P^*_{X'Y})]\}\big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big] \\
&\quad \le e^{-2n\varepsilon}. \qquad (168)
\end{aligned}$$

From (136) and (168), we obtain

$$\begin{aligned}
&\mathbb{P}\big[Z_{mm'}(\boldsymbol{y}) > (1-4\delta_n^2)e^{-2\delta_n} \\
&\quad \times \exp\{n[\alpha(R+2\varepsilon, P_Y) + \varepsilon]\}\big| \boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}\big] \\
&\dot{\le} e^{-2n\varepsilon} \qquad (169)
\end{aligned}$$

where the constant in $\dot{\le}$ does not depends on $\boldsymbol{x}_m, \boldsymbol{x}_{m'}$.

It follows from (128) and (169) that

$$\begin{aligned}
&\mathbb{P}\big[(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'}), Z_{mm'}(\boldsymbol{y}) > (1-4\delta_n^2)e^{-2\delta_n} \\
&\quad \times \exp\{n[\alpha(R+2\varepsilon, P_Y) + \varepsilon]\}\big] \\
&\dot{\le} \sum_{\substack{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'}) \\ d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta}} \mathbb{P}(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) e^{-2n\varepsilon} \qquad (170) \\
&\le \sum_{\substack{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'}) \\ d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta}} \left(\frac{1}{1-e^{-n\delta}}\right)^2 \frac{1}{|\mathcal{T}(Q_X)|^2} e^{-2n\varepsilon} \qquad (171) \\
&\dot{\le} e^{-nI_P(X;X')} e^{-2n\varepsilon}, \qquad (172)
\end{aligned}$$

where (171) follows from Lemma 4. By combining (126) and (172), we obtain

$$\mathbb{P}\big[\mathcal{G}_n^c(P_{Y|XX'})\cap\mathcal{E}^c(P_{XX'})\big]\dot{\leq} e^{-2n\varepsilon}. \tag{173}$$

On the other hand, by Lemma 6, we also have

$$\Pr[\mathcal{E}^c(P_{XX'})]\to 1. \tag{174}$$

Now, for any fixed joint-type $P_{XX'}\in\mathcal{Q}(Q_X)$ such that $I_P(X;X')<2R-\varepsilon$, define

$$\mathcal{F}_n(P_{XX'})\triangleq\bigcap_{P_{Y|XX'}}\{\mathcal{G}_n(P_{Y|XX'})\cap\mathcal{E}^c(P_{XX'})\}. \tag{175}$$

Then, from (173) and (174), for any fixed joint-type $P_{XX'}\in\mathcal{Q}(Q_X)$ such that $I_P(X;X')<2R-\varepsilon$, we have

$$\begin{aligned}
&\mathbb{P}[\mathcal{F}_n^c(P_{XX'})]\\
&=\mathbb{P}\bigg[\bigcup_{P_{Y|XX'}}\{\mathcal{G}_n^c(P_{Y|XX'})\cap\mathcal{E}^c(P_{XX'})\}\cup\mathcal{E}(P_{XX'})\bigg] &(176)\\
&\leq\mathbb{P}\bigg[\bigcup_{P_{Y|XX'}}\{\mathcal{G}_n^c(P_{Y|XX'})\cap\mathcal{E}^c(P_{XX'})\}\bigg]+\mathbb{P}[\mathcal{E}(P_{XX'})] &(177)\\
&\leq\sum_{P_{Y|XX'}}\mathbb{P}\big[\mathcal{G}_n^c(P_{Y|XX'})\cap\mathcal{E}^c(P_{XX'})\big]+\mathbb{P}[\mathcal{E}(P_{XX'})] &(178)\\
&\dot{\leq}|\mathcal{T}(P_{Y|XX'})|e^{-2n\varepsilon}+o(1) &(179)\\
&\to 0, &(180)
\end{aligned}$$

which leads to $\mathbb{P}[\mathcal{F}_n(P_{XX'})]\to 1$ as $n\to\infty$.

Now, for a given code $\mathcal{C}_n\in\mathcal{F}_n(P_{XX'})$, define

$$\begin{aligned}
\mathcal{V}(\mathcal{C}_n,P_{Y|XX'})=\{(m,m',\boldsymbol{y}):Z_{mm'}(\boldsymbol{y})\\
\leq\exp[n(\alpha(R+2\varepsilon,P_Y)+\varepsilon)]\},
\end{aligned} \tag{181}$$

and

$$\mathcal{V}_{m,m'}(\mathcal{C}_n,P_{Y|XX'})=\{\boldsymbol{y}:(m,m',\boldsymbol{y})\in\mathcal{V}(\mathcal{C}_n,P_{Y|XX'})\}. \tag{182}$$

Then, by definition of $\mathcal{G}_n(P_{Y|XX'})$ in (122), for any fixed joint type $P_{XX'}\in\mathcal{Q}(Q_X)$ such that $I_P(X;X')<2R-\varepsilon$ and $d(P_{XX'})>\Delta$, and for any $\mathcal{C}_n\in\mathcal{F}_n(P_{XX'})$, it holds that

$$\begin{aligned}
&\sum_{m,m'}\mathbb{1}\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\}\\
&\quad\times\frac{|\mathcal{T}(P_{Y|XX'})\cap\mathcal{V}_{m,m'}(\mathcal{C}_n,P_{Y|XX'})|}{|\mathcal{T}(P_{Y|XX'})|}\\
&\quad\geq(1-4\delta_n^2)e^{-2\delta_n}\exp\big[n(2R-I_P(X;X')-3\varepsilon/2)\big].
\end{aligned} \tag{183}$$

Now, let

$$\begin{aligned}
P_{XX'}^*:=\mathop{\arg\min}_{\substack{P_{XX'}:P_{X'}=P_X,\\ I_P(X;X')\leq 2R,d(P_{XX'})>\Delta}}\big\{\Gamma(P_{XX'},R)\\
+I_P(X;X')-R\big\}.
\end{aligned} \tag{184}$$

Then, for any $\rho>1$, we have

$$\begin{aligned}
&\mathbb{E}[(P_{\mathrm{e}}(\mathcal{C}_n))^{1/\rho}]\\
&=\mathbb{E}\Bigg[\Bigg(\frac{1}{M}\sum_m\sum_{m'\neq m}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{X}_m)\\
&\quad\times\frac{\exp\{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})\}+Z_{mm'}(\boldsymbol{y})}\Bigg)^{1/\rho}\Bigg] &(185)\\
&=\sum_{\mathcal{C}_n}\mathbb{P}[\mathcal{C}_n]\Bigg(\frac{1}{M}\sum_m\sum_{m'\neq m}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x}_m)\\
&\quad\times\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}+Z_{mm'}(\boldsymbol{y})}\Bigg)^{1/\rho} &(186)\\
&=\sum_{\mathcal{C}_n}\mathbb{P}[\mathcal{C}_n]\Bigg(\frac{1}{M}\sum_{P_{XX'}\in\mathcal{Q}(Q_X)}\sum_m\sum_{m'\neq m}\\
&\quad\mathbb{1}\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\}\\
&\quad\times\sum_{P_{Y|XX'}}\sum_{\boldsymbol{y}\in\mathcal{T}(P_{Y|XX'})}W(\boldsymbol{y}|\boldsymbol{x}_m)\\
&\quad\times\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}+Z_{mm'}(\boldsymbol{y})}\Bigg)^{1/\rho} &(187)\\
&=\sum_{\mathcal{C}_n}\mathbb{P}[\mathcal{C}_n]\Bigg(\frac{1}{M}\sum_{P_{XX'}\in\mathcal{Q}(Q_X)}\sum_m\sum_{m'\neq m}\\
&\quad\mathbb{1}\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\}\\
&\quad\times\sum_{P_{Y|XX'}}\sum_{\boldsymbol{y}\in\mathcal{T}(P_{Y|XX'})}W(\boldsymbol{y}|\boldsymbol{x}_m)\\
&\quad\times\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}+Z_{mm'}(\boldsymbol{y})}\Bigg)^{1/\rho} &(188)\\
&\geq\sum_{\mathcal{C}_n\in\mathcal{F}_n(P_{XX'}^*)}\mathbb{P}[\mathcal{C}_n]\Bigg(\frac{1}{M}\sum_m\sum_{m'\neq m}\\
&\quad\mathbb{1}\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'}^*)\}\\
&\quad\times\sum_{P_{Y|XX'}}\sum_{\boldsymbol{y}\in\mathcal{T}(P_{Y|XX'})\cap\mathcal{V}_{m,m'}(\mathcal{C}_n,P_{Y|XX'})}W(\boldsymbol{y}|\boldsymbol{x}_m)\\
&\quad\times\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}+Z_{mm'}(\boldsymbol{y})}\Bigg)^{1/\rho} &(189)\\
&\doteq\sum_{\mathcal{C}_n\in\mathcal{F}_n(P_{XX'}^*)}\mathbb{P}[\mathcal{C}_n]\Bigg(\frac{1}{M}\sum_{P_{Y|XX'}}\sum_m\sum_{m'\neq m}\\
&\quad\mathbb{1}\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'}^*)\}\\
&\quad\times\frac{|\mathcal{T}(P_{Y|XX'})\cap\mathcal{V}_{m,m'}(\mathcal{C}_n,P_{Y|XX'})|}{|\mathcal{T}(P_{Y|XX'})|}\\
&\quad\times\exp\bigg\{-n[D(P_{Y|X}\|W|Q_X)]+I_P(X';Y|X)\\
&\quad+[\max\{g(P_{XY}),\alpha(R+2\varepsilon,P_Y)+\varepsilon\}-g(P_{X'Y})]_+\bigg\}\Bigg)^{1/\rho} &(190)
\end{aligned}$$

$$\geq \sum_{\mathcal{C}_n \in \mathcal{F}_n(P^*_{XX'})} \mathbb{P}[\mathcal{C}_n]\Bigg(\frac{1}{M}\sum_{P_{Y|XX'}}(1-4\delta_n^2)e^{-2\delta_n} \times \exp\left[n(2R - I_{P^*}(X;X') - 3\varepsilon/2)\right] \times \exp\bigg\{-n[D(P_{Y|X}\|W|Q_X)] + I_P(X';Y|X) + [\max\{g(P_{XY}),\alpha(R+2\varepsilon,P_Y)+\varepsilon\} - g(P_{X'Y})]_+\bigg\}\Bigg)^{1/\rho} \tag{191}$$

$$\doteq \mathbb{P}[\mathcal{F}_n^c(P^*_{XX'})]\Bigg(\sum_{P_{Y|XX'}}(1-4\delta_n^2)e^{-2\delta_n} \times \exp\left[n(R - I_{P^*}(X;X') - 3\varepsilon/2)\right] \times \exp\bigg\{-n[D(P_{Y|X}\|W|Q_X)] + I_P(X';Y|X) + [\max\{g(P_{XY}),\alpha(R+2\varepsilon,P_Y)+\varepsilon\} - g(P_{X'Y})]_+\bigg\}\Bigg)^{1/\rho} \tag{192}$$

$$\doteq \mathbb{P}[\mathcal{F}_n^c(P^*_{XX'})]\Bigg(\exp\left[n(R - I_{P^*}(X;X') - 3\varepsilon/2)\right] \times \exp[-n\Gamma(P_{XX'},R+2\varepsilon)]\Bigg)^{1/\rho}, \tag{193}$$

where (188) follows from Tonelli's theorem [30], (190) follows from (169), and (191) follows from (183), (193) follows from $\delta_n \to 0$ and the definition of $\Gamma(P_{XX'},R)$.

From (193), it follows that

$$\begin{aligned} &E^{\mathrm{rgv}}_{\mathrm{trc}}(R,Q_X,d,\Delta) \\ &\quad= -\frac{1}{n}\lim_{\rho\to\infty}\rho\log\left(\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)^{1/\rho}]\right) \\ &\quad\leq \Gamma(P^*_{XX'},R) + I_{P^*}(X;X') - R + O(\varepsilon) \qquad (194)\\ &\quad= \min_{\substack{P_{XX'}:P_{X'}=P_X,\\ I_P(X;X')\leq 2R, d(P_{XX'})>\Delta}} \Big\{\Gamma(P_{XX'},R) + I_P(X;X') - R\Big\} + O(\varepsilon) \qquad (195)\end{aligned}$$

for any $\varepsilon > 0$. By taking $\varepsilon \to 0$, we obtain (120). This concludes the proof of Proposition 2. ■

V. Concentration Properties

In this section, we study the concentration properties of the RGV ensemble with GLD. In particular, we study the lower tail $\mathbb{P}\big[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \leq E_0\big]$ and derive both upper and lower bounds. We show that both bounds exhibit an exponential decay. We also derive upper and lower bounds to the upper tail $\mathbb{P}\big[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\big]$. We show that the upper tail exhibits a doubly-exponential behavior.

A. Lower Tail

In this section, we derive exponential upper and lower bounds to the lower tail probability. Before proceeding, we define the following sets

$$\mathcal{L}(R,E_0) \triangleq \{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \; [2R - I_P(X;X')]_+ \geq \Gamma(P_{XX'},R) + R - E_0\}, \tag{196}$$

$$\mathcal{M}(R,E_0) \triangleq \{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \; [2R - I_P(X;X')]_+ \geq \Lambda(P_{XX'},R) + R - E_0\} \tag{197}$$

where

$$\Lambda(P_{XX'},R) = \min_{P_{Y|XX'}}\{D(P_{Y|X}\|W|Q_X) + I_P(X';Y|X) + \beta(R,P_Y) - g(P_{X'Y})\}, \tag{198}$$

$$\beta(R,P_Y) = \max_{P_{\tilde{X}|Y}:P_{\tilde{X}}=Q_X}\{g(P_{\tilde{X}Y}) + [R - I_P(\tilde{X};Y)]_+\}. \tag{199}$$

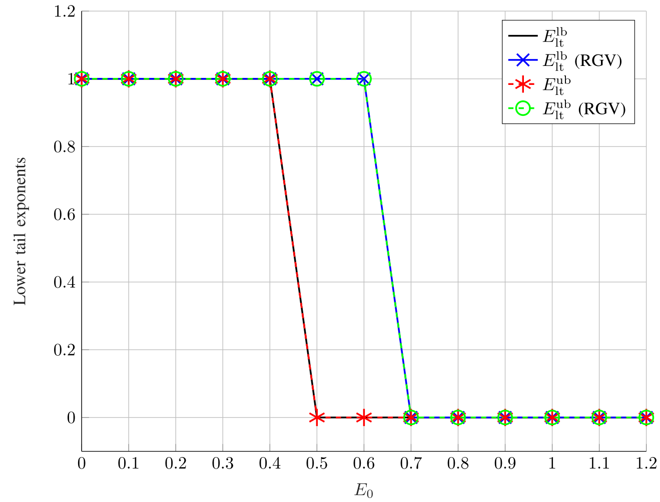

Fig. 2. Lower tail exponents for constant composition and RGV codes for the Z-channel.

We have the following result.

Theorem 2: Consider the ensemble of RGV codes $\mathcal{C}_n$ of rate R and composition Q_X satisfying condition (28). Then, it holds that

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \leq E_0\right] \dot{\leq} \exp\{-nE^{\mathrm{ub}}_{\mathrm{lt}}(R,E_0)\}, \tag{200}$$

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \leq E_0\right] \dot{\geq} \exp\{-nE^{\mathrm{lb}}_{\mathrm{lt}}(R,E_0)\}. \tag{201}$$

where

$$E^{\mathrm{ub}}_{\mathrm{lt}}(R,E_0) \triangleq \min_{P_{XX'}\in\mathcal{L}(R,E_0)}[I_P(X;X') - 2R]_+, \tag{202}$$

$$E^{\mathrm{lb}}_{\mathrm{lt}}(R,E_0) \triangleq \min_{P_{XX'}\in\mathcal{M}(R,E_0)}[I_P(X;X') - 2R]_+, \tag{203}$$

respectively.

Before proceeding with the proof, we discuss an example in Figure 2 where the lower tail bounds are shown for the Z-channel with crossover probability $w = 0.001$ and $R = 0.2$. In particular, we show the lower tail upper and lower bounds on the tail exponent for constant composition and for the RGV ensemble with $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -R$. The numerical results show that $E^{\mathrm{ub}}_{\mathrm{lt}} = E^{\mathrm{lb}}_{\mathrm{lt}}$ for the both constant composition and RGV ensembles. This can be explained by the fact that there is only one empirical channel $P_{X'Y}$ for each output type P_Y for this case [20, p. 5046]. Hence, $[\max\{g(P_{XY}),\alpha(R,P_Y)\} - g(P_{X'Y})] = [R - I(q)]_+ = \beta(R,P_Y) - g(P_{X'Y})$, which leads to $\Lambda = \Gamma$ for any R and crossover probability. Fig. 2 illustrates that the lower tail for the RGV code ensemble decays faster than that for the constant composition ensemble. This can be explained by the the fact that at $R = 0.2$ the typical error exponent of the RGV ensemble is higher than that for constant composition (see Figure 1).

1) Proof of the Lower Tail Upper Bound: Let

$$\mathcal{B}_\varepsilon(m,\boldsymbol{y}) = \Big\{\mathcal{C}_n : Z_m(\boldsymbol{y}) \le \exp\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\}\Big\}, \quad (204)$$

and

$$\mathcal{B}_\varepsilon \triangleq \bigcup_{m=1}^{M}\bigcup_{\boldsymbol{y}}\mathcal{B}_\varepsilon(m,\boldsymbol{y}). \quad (205)$$

Then, under the condition (28), by Lemma 14, we have

$$\mathbb{P}\big\{\mathcal{B}_\varepsilon(m,\boldsymbol{y})\big\} \le \exp\bigg\{-e^{n\varepsilon}\bigg[1 - \frac{e^{-n(\varepsilon+\delta)}}{1-e^{-n\delta}} - e^{-n\varepsilon}(1+n\varepsilon)\bigg]\bigg\}. \quad (206)$$

Hence, by the union bound, we have

$$\begin{aligned}\mathbb{P}\{\mathcal{B}_\varepsilon\} &\le \sum_{m=1}^{M}\sum_{\boldsymbol{y}}\mathbb{P}\big\{\mathcal{B}_\varepsilon(m,\boldsymbol{y})\big\} &(207)\\ &\le \sum_{m=1}^{M}\sum_{\boldsymbol{y}}\exp\bigg\{-e^{n\varepsilon}\bigg[1-\frac{e^{-n(\varepsilon+\delta)}}{1-e^{-n\delta}} - e^{-n\varepsilon}(1+n\varepsilon)\bigg]\bigg\} &(208)\\ &\le e^{nR}|\mathcal{Y}|^n\exp\bigg\{-e^{n\varepsilon}\bigg[1-\frac{e^{-n(\varepsilon+\delta)}}{1-e^{-n\delta}} - e^{-n\varepsilon}(1+n\varepsilon)\bigg]\bigg\} &(209)\end{aligned}$$

where (208) follows from (206), which decays double-exponentially fast.

Now, by using the same arguments as [15, Proof of Theorem 1], we have

$$\begin{aligned}&\mathbb{P}\bigg[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \le E_0\bigg]\\ &\le \mathbb{P}\bigg[\mathcal{C}_n\in\mathcal{B}_\varepsilon^c, \frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m} e^{-n\Gamma(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}},R-\varepsilon)} \ge e^{-nE_0}\bigg]\\ &\quad + \mathbb{P}\{\mathcal{B}_\varepsilon\} &(210)\\ &\dot{\le}\ \mathbb{P}\bigg[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m} e^{-n\Gamma(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}},R-\varepsilon)} \ge e^{-nE_0}\bigg] &(211)\\ &= \mathbb{P}\bigg[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m} e^{-n\Gamma(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}},R-\varepsilon)}\\ &\quad\times \mathbb{1}\big\{d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta\big\} \ge e^{-nE_0}\bigg] &(212)\\ &= \mathbb{P}\bigg[\sum_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta} N(P_{XX'})\\ &\quad\times\exp\big\{-n\Gamma(P_{XX'},R-\varepsilon)\big\} \ge e^{n(R-E_0)}\bigg] &(213)\\ &\doteq \max_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta}\mathbb{P}\bigg[N(P_{XX'}) \ge\\ &\quad \exp\{n(\Gamma(P_{XX'},R-\varepsilon)+R-E_0)\}\bigg] &(214)\end{aligned}$$

where (210) follows from [15, Eq. (60)], and (212) follows from the fact that all codes $\mathcal{C}_n$ in the RGV ensemble satisfy $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'}) > \Delta$ for all $m \neq m'$.

Now, define

$$\mathcal{S}_\varepsilon(R,E_0) \triangleq \Big\{P_{XX'}\in\mathcal{Q}(Q_X) : [2R - I_P(X;X')]_+ \ge \Gamma(P_{XX'},R-\varepsilon) + R - E_0\Big\}. \quad (215)$$

Then, from (214) and Lemma 10, under the condition 28, we obtain

$$\mathbb{P}\bigg[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \le E_0\bigg] \dot{\le} e^{-nE_{\mathrm{lt}}^{\mathrm{ub}}(R,E_0,\varepsilon)}, \quad (216)$$

where

$$\begin{aligned}&E_{\mathrm{lt}}^{\mathrm{ub}}(R,E_0,\varepsilon)\\ &\triangleq \min_{P_{XX'}:d(P_{XX'})>\Delta}\begin{cases}[I_P(X;X')-2R]_+, & P_{XX'}\in\mathcal{S}_\varepsilon(R,E_0)\\ +\infty, & \text{otherwise}\end{cases} &(217)\\ &= \min_{P_{XX'}\in\mathcal{S}_\varepsilon(R,E_0):d(P_{XX'})>\Delta}[I_P(X;X')-2R]_+, &(218)\end{aligned}$$

with the convention that the minimum over an empty set is defined as infinity. Since ε can take any positive value, from (216) and (218), we obtain (201). This concludes our proof of the upper bound in Theorem 2.

2) Proof of the Lower Tail Lower Bound: The proof follows similar arguments as [15, Section B]. For the RGV ensemble, however, existing techniques to lower bound on the probability of the lower tail for the constant composition codes cannot be applied. For example, due to the dependence among codewords, key proposition [15, Prep. 4] can no longer be applied. We develop new techniques to deal with the dependence among codewords.

For a given $(m,m')\in[M]_*^2$, and $\boldsymbol{y}\in\mathcal{Y}^n$, define

$$Z_{m,m'}(\boldsymbol{y}) = \sum_{\tilde{m}\in\{1,2,\cdots,M\}\setminus\{m,m'\}}\exp\big\{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}},\boldsymbol{y}})\big\}. \quad (219)$$

Let $\sigma > 0$ and define the set

$$\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y}) = \Big\{C_n : Z_{mm'}(\boldsymbol{y}) \ge \exp\{n(\beta(R,\hat{P}_{\boldsymbol{y}})+\sigma)\}\Big\}, \quad (220)$$

and its complement $\hat{\mathcal{G}}_n(\sigma,m,m',\boldsymbol{y}) = \hat{\mathcal{B}}_n^c(\sigma,m,m',\boldsymbol{y})$, where $\beta(R,P_Y)$ is defined in (199). Let

$$\hat{\mathcal{B}}_n(\sigma) = \bigcup_{m=1}^{M}\bigcup_{m'\neq m}\bigcup_{\boldsymbol{y}}\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y}), \quad (221)$$

and

$$\hat{\mathcal{G}}_n(\sigma) = \hat{\mathcal{B}}_n^c(\sigma). \quad (222)$$

Let $\varepsilon > 0$ and define

$$\tilde{\Lambda}(P_{XX'}, R, \varepsilon) = \min_{P_{Y|XX'}} \{D(P_{Y|X} \| W|Q_X) + I_P(X';Y|X) + [\max\{g(P_{XY}), \beta(R, P_Y) + \varepsilon\} - g(P_{X'Y})]_+\}. \tag{223}$$

Then, we have

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \le E_0\right] \dot{\ge} \mathbb{P}\left[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\varepsilon), \frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m} e^{-n\tilde{\Lambda}(\hat{P}_{\boldsymbol{X}_{m'},\boldsymbol{X}_m}, R, \varepsilon)} \ge e^{-nE_0}\right] \tag{224}$$

$$= \mathbb{P}\left[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\varepsilon), \frac{1}{M}\sum_{m=1}^{M}\sum_{\substack{m'\neq m\\ d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta}} e^{-n\tilde{\Lambda}(\hat{P}_{\boldsymbol{X}_{m'},\boldsymbol{X}_m}, R, \varepsilon)} \ge e^{-nE_0}\right], \tag{225}$$

where (224) follows from [15, Eq. (83)], and (225) follows from the fact that $d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) > \Delta$ for any RGV code.

On the other hand, we also have

$$\tilde{\Lambda}(P_{XX'}, R, \varepsilon) = \Lambda(P_{XX'}, R) + \varepsilon. \tag{226}$$

Hence, from (225) and (226), we obtain

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \le E_0\right] \dot{\ge} \mathbb{P}\big[\hat{\mathcal{G}}_n(\varepsilon) \cap \mathcal{G}_0\big], \tag{227}$$

where

$$\mathcal{G}_0 = \left\{\mathcal{C}_n : \sum_{m=1}^{M}\sum_{\substack{m'\neq m\\ d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta}} e^{-n\tilde{\Lambda}(\hat{P}_{\boldsymbol{X}_{m'},\boldsymbol{X}_m}, R, \varepsilon)} \ge e^{n(R-E_0)}\right\}. \tag{228}$$

It then follows that

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \le E_0\right] \dot{\ge} \mathbb{P}\big[\hat{\mathcal{G}}_n(\varepsilon) \cap \mathcal{G}_0\big] \tag{229}$$

$$= \mathbb{P}[\mathcal{G}_0] - \mathbb{P}\big[\mathcal{G}_0 \cap \hat{\mathcal{B}}_n(\varepsilon)\big] \tag{230}$$

$$\ge \mathbb{P}[\mathcal{G}_0] - \sum_{m=1}^{M}\sum_{m'\neq m}\sum_{\boldsymbol{y}} \mathbb{P}\big[\hat{\mathcal{B}}_n(\varepsilon, m, m', \boldsymbol{y}) \cap \mathcal{G}_0\big]. \tag{231}$$

Now, observe that

$$\mathbb{P}[\mathcal{G}_0] = \mathbb{P}\left[\sum_{\substack{P_{XX'}\in\mathcal{Q}(Q_X):\\ d(P_{XX'})>\Delta}} N(P_{XX'})e^{-n(\Lambda(P_{XX'},R)+\varepsilon)} \ge e^{n(R-E_0)}\right] \tag{232}$$

$$\doteq \sum_{\substack{P_{XX'}\in\mathcal{Q}(Q_X):\\ d(P_{XX'})>\Delta}} \mathbb{P}\left[N(P_{XX'}) \ge e^{n(\Lambda(P_{XX'},R)+R-E_0+\varepsilon)}\right]. \tag{233}$$

Define the set $\mathcal{S}'_\varepsilon(R, E_0) = \{P_{XX'} : [2R - I_P(X;X')]_+ \ge \Lambda(P_{XX'}, R) + R - E_0 + \varepsilon\}$.

Then, under condition (28), by Proposition 10, it holds that

$$\mathbb{P}[\mathcal{G}_0] \doteq \exp\{-nE_{\mathrm{lt}}^{\mathrm{lb}}(R, E_0, \varepsilon)\}, \tag{234}$$

where

$$E_{\mathrm{lt}}^{\mathrm{lb}}(R, E_0, \varepsilon) = \min_{\substack{P_{XX'}\in\mathcal{Q}(Q_X):\\ d(P_{XX'})>\Delta}} \begin{cases} [I_P(X;X') - 2R]_+ & P_{XX'} \in \mathcal{S}'_\varepsilon(R, E_0) \\ +\infty & P_{XX'} \notin S'_\varepsilon(R, E_0) \end{cases} \tag{235}$$

$$= \min_{\substack{P_{XX'}\in\{P_{XX'}\in\mathcal{Q}(Q_X):\\ d(P_{XX'})>\Delta\}\cap\mathcal{S}'_\varepsilon(R,E_0)}} [I_P(X;X') - 2R]_+. \tag{236}$$

Now, we study the second term in (231). For any joint type $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, define

$$N_{\boldsymbol{y}}(P_{XY}) \triangleq \sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\}. \tag{237}$$

Then, we have

$$\mathbb{P}\big[\hat{\mathcal{B}}_n(\varepsilon, \hat{m}, \ddot{m}, \boldsymbol{y}) \cap \mathcal{G}_0\big] = \mathbb{P}\left[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}},\boldsymbol{y}})} \ge e^{n(\beta(R,\hat{P}_{\boldsymbol{y}})+\varepsilon)}, \sum_{m=1}^{M}\sum_{m'\neq m} e^{-n(\Lambda(\hat{P}_{\boldsymbol{x}_{\tilde{m}},\boldsymbol{x}_m},R)+\varepsilon)} \ge e^{n(R-E_0)}\right] \tag{238}$$

$$\le \mathbb{P}\left[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}},\boldsymbol{y}})} \ge e^{n(\beta(R,\hat{P}_{\boldsymbol{y}})+\varepsilon)}\right] \tag{239}$$

$$\le \mathbb{P}\left[\sum_{P_{XY}:P_X=Q_X} N_{\boldsymbol{y}}(P_{XY})e^{ng(P_{XY})} \ge e^{n(\beta(R,\hat{P}_{\boldsymbol{y}})+\varepsilon)}\right] \tag{240}$$

$$\dot{\le} \sum_{P_{XY}:P_X=Q_X} \mathbb{P}\left[N_{\boldsymbol{y}}(P_{XY}) \ge e^{n(\beta(R,\hat{P}_{\boldsymbol{y}})-g(P_{XY})+\varepsilon)}\right] \tag{241}$$

$$\le \sum_{P_{XY}:P_X=Q_X} \mathbb{P}\left[N_{\boldsymbol{y}}(P_{XY}) \ge e^{n([R-I_P(X;Y)]_++\varepsilon)}\right] \tag{242}$$

$$= \sum_{P_{XY}:P_X=Q_X} \mathbb{P}\left[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \ge e^{n([R-I_P(X;Y)]_++\varepsilon)}\right] \tag{243}$$

$$= \sum_{P_{XY}:P_X=Q_X, I_P(X;Y)>0} \mathbb{P}\left[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \ge e^{n([R-I_P(X;Y)]_++\varepsilon)}\right], \tag{244}$$

where (242) follows from (199), and (244) follows from

$$\mathbb{P}\left[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\}\right.$$

$$\geq e^{n([R-I_P(X;Y)]_+ + \varepsilon)}\Big] = 0 \tag{245}$$

if $I_P(X;Y) = 0$.

Now, in order to bound $\mathbb{P}\big[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \geq e^{n([R-I_P(X;Y)]_+ + \varepsilon)}\big]$ for each joint type P_{XY} such that $P_X = Q_X$, we will use the following lemma.

Lemma 15: [31, Lemma 1.8] Suppose that $X_1, X_2, \cdots, X_n$ are random variables such that $0 \leq X_i \leq 1$, for $i = 1, 2, \cdots, n$. Set $p = \frac{1}{n}\sum_i \mathbb{E}[X_i]$ and fix a real number t such that $np + 1 < t < n$. If $\varepsilon_0 > 0$ is such that $t - 1 = np(1+\varepsilon_0)$, then

$$\mathbb{P}\left[\sum_{i=1}^n X_i \geq t\right] \leq 2\ e^{-nD(p(1+\varepsilon_0)\|p)}. \tag{246}$$

More specifically, for any $\tilde{m} \in [M] \setminus \{\hat{m}, \ddot{m}\}$, observe that

$$\begin{aligned} &\mathbb{E}\big[\mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\}\big] \\ &= \mathbb{P}\big[(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\big] &(247)\\ &= \sum_{\boldsymbol{x}_{\tilde{m}}\in\mathcal{T}(Q_X):(\boldsymbol{x}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{XY})} \mathbb{P}(\boldsymbol{x}_{\tilde{m}}) &(248)\\ &= \sum_{\boldsymbol{x}_{\tilde{m}}\in\mathcal{T}(Q_X):(\boldsymbol{x}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{XY})} \frac{1}{|\mathcal{T}(Q_X)|} &(249)\\ &\doteq e^{-nI_P(X;Y)}, &(250)\end{aligned}$$

where (249) follows from 3, and (250) follows from [28].

It follows from (250) that

$$\begin{aligned} p &\triangleq \frac{1}{M-2}\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{E}\big[\mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\}\big] &(251)\\ &\doteq e^{-nI_P(X;Y)}. &(252)\end{aligned}$$

Now, there exists a $\delta(\varepsilon) < \varepsilon$ such that $\min\{I_P(X;Y) : I_P(X;Y) > 0\} > \delta(\varepsilon)$. Then, we have

$$\begin{aligned} &\sum_{P_{XY}:P_X=Q_X, I_P(X;Y)>0} \mathbb{P}\Big[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \geq e^{n([R-I_P(X;Y)]_+ + \varepsilon)}\Big] \\ &\leq \sum_{P_{XY}:P_X=Q_X, I_P(X;Y)>0} \mathbb{P}\Big[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \geq e^{n([R-I_P(X;Y)]_+ + \delta(\varepsilon))}\Big]. &(253)\end{aligned}$$

By applying Lemma 15 for the sequence of Bernoulli random variables $\big\{\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\}\big\}_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}}$ with $t = e^{n([R-I_P(X;Y)]_+ + \delta(\varepsilon))}$, we obtain

$$\begin{aligned} &\mathbb{P}\Big[\sum_{\tilde{m}\in[M]\setminus\{\hat{m},\ddot{m}\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \geq e^{n([R-I_P(X;Y)]_+ + \delta(\varepsilon))}\Big] \\ &\overset{\circ}{\leq} \exp\Big\{-MD\big(e^{n([R-I_P(X;Y)]_+ - R + \delta(\varepsilon))}\|e^{-nI_P(X;Y)}\big)\Big\} &(254)\\ &\overset{\circ}{\leq} \exp\Big\{-e^{n([R-I_P(X;Y)]_+ + \delta(\varepsilon))}\Big\} &(255)\end{aligned}$$

for any joint type $P_{XY} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$ such that $P_X = Q_X$, where (255) follows from the fact that $D(a\|b) \geq a\big(\log\frac{a}{b} - 1\big)$ [29].

It follows from that

$$\begin{aligned} &\mathbb{P}\big[\hat{\mathcal{B}}_n(\varepsilon, \hat{m}, \ddot{m}, \boldsymbol{y}) \cap \mathcal{G}_0\big] \\ &\overset{\circ}{\leq} \max_{P_{XY}} \exp\Big\{-e^{n([R-I_P(X;Y)]_+ + \delta(\varepsilon))}\Big\} &(256)\\ &\overset{\circ}{\leq} \exp\big\{-e^{n\delta(\varepsilon)}\big\}. &(257)\end{aligned}$$

From (231), (234), and (257), we finally obtain

$$\begin{aligned} &\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \leq E_0\right] \\ &\dot{\geq} \exp\big\{-nE_{\mathrm{lt}}^{\mathrm{lb}}(R, E_0, \varepsilon)\big\} - \sum_{m=1}^M \sum_{m'\neq m} \sum_{\boldsymbol{y}} \exp\big\{-e^{n\delta(\varepsilon)}\big\} &(258)\\ &\doteq \exp\big\{-nE_{\mathrm{lt}}^{\mathrm{lb}}(R, E_0, \varepsilon)\big\} - e^{2nR}|\mathcal{Y}|^n \exp\big\{-e^{n\delta(\varepsilon)}\big\} &(259)\\ &\doteq \exp\big\{-nE_{\mathrm{lt}}^{\mathrm{lb}}(E, E_0, \varepsilon)\big\}. &(260)\end{aligned}$$

Due to the arbitrariness of $\varepsilon > 0$, it follows that

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \leq E_0\right] \dot{\geq} \exp\{-nE_{\mathrm{lt}}^{\mathrm{lb}}(R, E_0)\}, \tag{261}$$

which proves the lower bound of Theorem 2.

B. Upper Tail

In this section, we derive double-exponential upper and lower bounds to the upper tail probability. First, we introduce some new notation which will be used throughout this section. Recall the definitions of $\mathcal{A}_1$ and $\mathcal{A}_2$ in (65) and (66), respectively. Let

$$\begin{aligned} \mathcal{V}(R, E_0) = \big\{&P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \\ &I_P(X;X') \leq 2R, \Lambda(P_{XX'}, R) + I_P(X;X') - R \leq E_0\big\}, \end{aligned} \tag{262}$$

$$\begin{aligned} \mathcal{U}(R, E_0) = \big\{&P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \\ &I_P(X;X') \leq 2R, \Gamma(P_{XX'}, R) + I_P(X;X') - R \leq E_0\big\}. \end{aligned} \tag{263}$$

and

$$\begin{aligned} \mathcal{A}_3 = \big\{&P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, I_P(X;X') \leq 2R, \\ &\Gamma(P_{XX'}, R-\varepsilon) + I_P(X;X') - R > E_0 + \varepsilon\big\}. \end{aligned} \tag{264}$$

Theorem 3: Consider the RGV ensemble $\mathcal{C}_n$ of rate R and composition Q_X satisfying condition (28). Assume that the conditions in Lemma 11 hold for $\mathcal{D} = \tilde{\mathcal{V}}(R, E_0, \sigma)$. Then, the upper tail can be bounded as

$$\mathbb{P}\Big[-\frac{1}{n} \log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0 \Big] \overset{\circ}{\leq} \exp\big\{ -\exp\big\{ n E_{\mathrm{ut}}^{\mathrm{ub}}(R, E_0) \big\} \big\} \tag{265}$$

where

$$E_{\mathrm{ut}}^{\mathrm{ub}}(R, E_0) = \max_{P_{XX'} \in \mathcal{V}(R,E_0)} \min \big\{ 2R - I_P(X;X'), \\ E_0 - \Lambda(P_{XX'}, R) - I_P(X;X') + R, R \big\}. \tag{266}$$

In addition, under the conditions

$$\max_{P_{XX'} \in \mathcal{A}_3} I_P(X;X') \leq \min_{P_{XX'} \in \mathcal{A}_2} I_P(X;X') \tag{267}$$

$$\min_{P_{XX'}: d(P_{XX'}) \leq \Delta} I_P(X;X') \geq \max_{P_{XX'}: d(P_{XX'}) > \Delta} I_P(X;X'), \tag{268}$$

we have that

$$\mathbb{P}\Big[-\frac{1}{n} \log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0 \Big] \overset{\circ}{\geq} \exp\big\{ -\exp\big\{ n E_{\mathrm{ut}}^{\mathrm{lb}}(R, E_0) \big\} \big\} \tag{269}$$

for all $E_0 < E_{\mathrm{ex}}(R, Q_X)$, where

$$E_{\mathrm{ut}}^{\mathrm{lt}}(R, E_0) = \max_{P_{XX'} \in \mathcal{U}(R,E_0)} \big\{ 2R - I_P(X;X') \big\}. \tag{270}$$

In Figure 3 we show the double-exponential bounds for the upper tail for constant composition and the RGV ensemble with $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -R$ for $R = 0.2$. We observe that for constant composition the decay is indeed double-exponential even if the bounds only coincide for high values of E_0 (above the TRC exponent). Instead, for the RGV ensemble, the bound $E_{\mathrm{ut}}^{\mathrm{lt}}(R, E_0) = 0$ for values of E_0 of interest. This implies that the decay of the upper tail for $E_{\mathrm{trc}}^{\mathrm{cc}} \leq E_0 \leq E_{\mathrm{ex}}^{\mathrm{cc}}$ is sub-double-exponential; for $E_0 > E_{\mathrm{ex}}$ the behavior of the upper tail is double-exponential as suggested by $E_{\mathrm{ut}}^{\mathrm{ut}}$ for the RGV ensemble. Figure. 3 also shows that the decay rate of RGV code is slower than the constant composition code. This can be explained by the the fact that the error probability in RGV code is expected to be smaller than the constant composition codes since the later is more structured as in the Fig. 2.

1) Proof of the Upper Tail Upper Bound: The proof is based on [15, Proof of Theorem 2] with important changes to account for the dependency among codewords in the RGV codebook ensemble. See also the proofs of Lemma 11 and Lemma 16 below for specific changes.

Lemma 16: For every $\sigma > 0$, under condition (28) the following holds

$$\mathbb{P}\{\hat{\mathcal{B}}_n(\sigma)\} \overset{\circ}{\leq} \exp\{-e^{n\sigma}\} \tag{271}$$

where $\hat{\mathcal{B}}_n(\sigma)$ has been defined in (221).

Proof: See Appendix L. ■

We start by defining the following set

$$\tilde{\mathcal{V}}(R, E_0, \sigma) \\ \triangleq \big\{ P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, I_P(X;X') \leq 2R, \\ \tilde{\Lambda}(P_{XX'}, R, \sigma) + I_P(X;X') - R \leq E_0 - \varepsilon \big\} \tag{272}$$

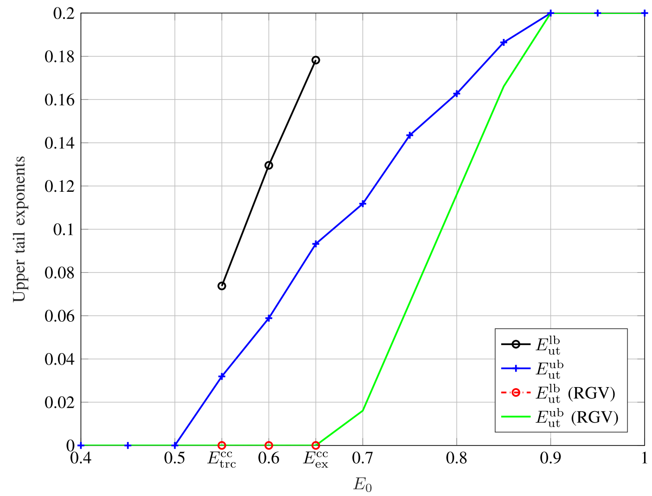

Fig. 3. Upper tail exponents for constant composition and RGV codes for the Z-channel.

for $\sigma > 0, \varepsilon > 0$, where $\tilde{\Lambda}(P_{XX'}, R, \varepsilon)$ was defined in (223).

Under condition (28), we have that

$$\mathbb{E}[N(P_{XX'})] \\ = \mathbb{E}\Bigg[\sum_{(m,m') \in [M]_*^2} \mathbb{1}\{ (\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'}) \} \Bigg] \tag{273}$$

$$= \sum_{(m,m') \in [M]_*^2} \mathbb{P}\big[(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'}) \big] \tag{274}$$

$$= \sum_{(m,m') \in [M]_*^2} \sum_{\boldsymbol{x}_m, \boldsymbol{x}_{m'} \in \mathcal{T}(P_{XX'})} \mathbb{P}\big(\boldsymbol{x}_m, \boldsymbol{x}_{m'} \big) \tag{275}$$

$$\doteq e^{n(2R - I_P(X;X'))}, \tag{276}$$

where (276) follows from Lemma 2.

For a given message pair $m, m' \in [M]_*^2$, and $\boldsymbol{y} \in \mathcal{Y}^n$, recall the definitions of $Z_{m,m'}(\boldsymbol{y})$, $\hat{\mathcal{B}}_n(\sigma)$, and $\hat{\mathcal{G}}_n(\sigma)$ in (219), (221), and (222), respectively. Then, we have

$$\mathbb{P}\Big[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n} \log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0 \Big] \tag{277}$$

$$\leq \mathbb{P}\Bigg[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), \frac{1}{M} \sum_{m=1}^{M} \sum_{m' \neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{X}_m) \\ \times \frac{e^{n g(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})}}{e^{n g(\hat{P}_{\boldsymbol{X}_m \boldsymbol{y}})} + e^{n g(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})} + Z_{mm'}(\boldsymbol{y})} \leq e^{-nE_0} \Bigg] \tag{278}$$

$$= \mathbb{P}\Bigg[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), \frac{1}{M} \sum_{m=1}^{M} \sum_{m' \neq m : d(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) > \Delta} \sum_{\boldsymbol{y}} \\ W(\boldsymbol{y}|\boldsymbol{X}_m) \frac{e^{n g(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})}}{e^{n g(\hat{P}_{\boldsymbol{X}_m \boldsymbol{y}})} + e^{n g(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})} + Z_{mm'}(\boldsymbol{y})} \leq e^{-nE_0} \Bigg] \tag{279}$$

$$\overset{\circ}{\leq} \min_{P_{XX'} \in \tilde{\mathcal{V}}(R, E_0, \sigma)} \mathbb{P}\Big[N(P_{XX'}) \leq e^{n(\tilde{\Lambda}(P_{XX'}, R, \sigma) + R - E_0)} \Big] \tag{280}$$

$$\overset{\circ}{\leq} \min_{P_{XX'}\in\tilde{\mathcal{V}}(R,E_0,\sigma)} \mathbb{P}\Big[N(P_{XX'}) \leq e^{n\left(2R-I_P(X;X')-\varepsilon\right)}\Big] \tag{281}$$

$$\overset{\circ}{\leq} \min_{p_{XX'}\in\tilde{\mathcal{V}}(R,E_0,\sigma)} \exp\big\{-\min\big(e^{n(2R-I_P(X;X'))}, e^{nR}\big)\big\} \tag{282}$$

where (278) follows from (11), (279) follows from the fact that $d(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) > \Delta$ with probability 1 by the RGV random codebook generation, (280) follows the same arguments to achieve [15, Eq. (146)], (281) follows from (272), and (282) follows from (276) and Lemma 11.

It follows from (282) that for $\sigma > 0$,

$$\mathbb{P}\Big[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big\} \overset{\circ}{\leq} \exp\big\{-\exp\big\{nE_1(R,E_0,\sigma)\big\}\big\}, \tag{283}$$

where

$$E_1(R,E_0,\sigma) = \max_{P_{XX'}\in\tilde{\mathcal{V}}(R,E_0,\sigma)} \min\{2R - I_P(X;X'), R\}. \tag{284}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{P}\Big[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \\
&= \mathbb{P}\Big[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \\
&\quad + \mathbb{P}\Big[\mathcal{C}_n \in \hat{\mathcal{G}}_n^c(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \qquad (285)\\
&= \mathbb{P}\Big[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \\
&\quad + \mathbb{P}\Big[\mathcal{C}_n \in \hat{\mathcal{B}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \qquad (286)\\
&\leq \mathbb{P}\Big[\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \\
&\quad + \mathbb{P}\big[\mathcal{C}_n \in \hat{\mathcal{B}}_n(\sigma)\big] \qquad (287)\\
&\overset{\circ}{\leq} \exp\big\{-\exp\big\{nE_1(R,E_0,\sigma)\big\}\big\} + \exp\{-e^{n\sigma}\} \qquad (288)
\end{aligned}$$

where (286) follows from $\hat{\mathcal{B}}_n(\sigma) = \hat{\mathcal{G}}_n^c(\sigma)$, (288) follows from Lemma 16 and (283).

Finally, by using the same arguments as to obtain [15, Eq. (175)] from [15, Eq. (153)], from (288), we obtain

$$\mathbb{P}\Big[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \overset{\circ}{\leq} \exp\Big\{-e^{nE_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0)}\Big\}, \tag{289}$$

which concludes our proof of the upper bound on the upper tail.

2) Proof of the Upper Tail Lower Bound: Let

$$\mathcal{B}_\varepsilon(m,\boldsymbol{y}) = \Big\{\mathcal{C}_n : Z_m(\boldsymbol{y}) \leq \exp\{n\alpha(R-\varepsilon, \hat{P}_{\boldsymbol{y}})\}\Big\}, \tag{290}$$

and

$$\mathcal{B}_\varepsilon \triangleq \bigcup_{m=1}^{M}\bigcup_{\boldsymbol{y}} \mathcal{B}_\varepsilon(m,\boldsymbol{y}). \tag{291}$$

Then, under condition (28), by Lemma 14 and the union bound, we have

$$\begin{aligned}
&\mathbb{P}\{\mathcal{B}_\varepsilon\} \\
&\leq e^{nR}|\mathcal{Y}|^n \exp\Big\{-e^{n\varepsilon}\Big[1 - \frac{e^{-n(\varepsilon+\delta)}}{1-e^{-n\delta}} - e^{-n\varepsilon}(1+n\varepsilon)\Big]\Big\}. \qquad (292)
\end{aligned}$$

Now, define $\mathcal{G}_\varepsilon(m,\boldsymbol{y}) = \mathcal{B}_\varepsilon^c(m,\boldsymbol{y})$ and $\mathcal{G}_\varepsilon = \mathcal{B}_\varepsilon^c$.

Recall the definition of $Z_m(\boldsymbol{y})$ in (98). We have that

$$\begin{aligned}
&\mathbb{P}\Big[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\Big] \\
&= \mathbb{P}\Big[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{X}_m) \\
&\qquad \times \frac{\exp\{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\} + Z_m(\boldsymbol{y})} \leq e^{-nE_0}\Big] \qquad (293)\\
&= \mathbb{P}\Big[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m: d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{X}_m) \\
&\qquad \times \frac{\exp\{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\} + Z_m(\boldsymbol{y})} \leq e^{-nE_0}\Big] \qquad (294)\\
&\geq \mathbb{P}\Big[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m: d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{X}_m) \\
&\qquad \times \frac{\exp\{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\} + Z_m(\boldsymbol{y})} \leq e^{-nE_0}, \mathcal{C}_n \in \mathcal{G}_\varepsilon\Big] \qquad (295)\\
&\overset{\circ}{\geq} \mathbb{P}\Big[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m: d(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}})>\Delta} \\
&\qquad \exp\big\{-n\Gamma(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}}, R-\varepsilon)\big\} \leq e^{-nE_0}, \mathcal{C}_n \in \mathcal{G}_\varepsilon\Big], \qquad (296)
\end{aligned}$$

where (294) follows from the fact that $\min_{i\neq j} d(\boldsymbol{x}_i, \boldsymbol{x}_j) > \Delta$ for all RGV code $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M)$, and (296) follows from the same arguments to obtain [15, Eq. (178)].

Now, define

$$\mathcal{E}_0 \triangleq \Big\{\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m: d(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}})>\Delta} \exp\big\{-n\Gamma(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}}, R-\varepsilon)\big\} \leq e^{-nE_0}\Big\}. \tag{297}$$

Then, we have

$$\begin{aligned}
&\mathbb{P}\Big[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m: d(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}})>\Delta} \\
&\qquad \exp\big\{-n\Gamma(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}}, R-\varepsilon)\big\} \leq e^{-nE_0}, \mathcal{C}_n \in \mathcal{G}_\varepsilon\Big] \\
&= \mathbb{P}\Big[\mathcal{C}_n \in \mathcal{E}_0, \mathcal{C}_n \in \mathcal{G}_\varepsilon\Big] \qquad (298)
\end{aligned}$$

$$= \mathbb{P}\bigg[\bigcap_{\tilde{m}=1}^{M}\bigcap_{\boldsymbol{y}}\mathcal{G}_{\varepsilon}(\tilde{m},\boldsymbol{y})\big|\mathcal{E}_0\bigg]\mathbb{P}(\mathcal{E}_0) \tag{299}$$

$$= \bigg(1-\mathbb{P}\bigg[\bigcup_{\tilde{m}=1}^{M}\bigcup_{\boldsymbol{y}}\mathcal{G}_{\varepsilon}^c(\tilde{m},\boldsymbol{y})\big|\mathcal{E}_0\bigg]\bigg)\mathbb{P}[\mathcal{E}_0] \tag{300}$$

$$\geq \bigg(1-\sum_{m=1}^{M}\sum_{\boldsymbol{y}}\mathbb{P}\big[\mathcal{G}_{\varepsilon}^c(\tilde{m},\boldsymbol{y})\big|\mathcal{E}_0\big]\bigg)\mathbb{P}[\mathcal{E}_0] \tag{301}$$

$$= \mathbb{P}[\mathcal{E}_0]-\sum_{m=1}^{M}\sum_{\boldsymbol{y}}\mathbb{P}\big[\mathcal{B}_{\varepsilon}(\tilde{m},\boldsymbol{y})\cap\mathcal{E}_0\big]. \tag{302}$$

Now, observe that

$$\mathbb{P}[\mathcal{E}_0] = \mathbb{P}\bigg[\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m: d(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{x}_{m'}})>\Delta} \exp\big\{-n\Gamma(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{x}_{m'}},R-\varepsilon)\big\} \leq e^{-nE_0}\bigg] \tag{303}$$

$$\doteq \mathbb{P}\bigg[\bigcap_{P_{XX'}\in\mathcal{Q}(Q_X): d(P_{XX'})>\Delta}\Big\{N(P_{XX'}) \leq e^{n(\Gamma(P_{XX'},R-\varepsilon)+R-E_0)}\Big\}\bigg], \tag{304}$$

where (304) follows by using the same arguments to achieve [15, Eq. (187)].

Recall the definition of $\mathcal{F}_0$ in (67) in Lemma 12, i.e.,

$$\mathcal{F}_0 = \bigcap_{P_{XX'}\in\mathcal{A}_1\cup\mathcal{A}_2}\big\{N(P_{XX'})=0\big\}. \tag{305}$$

Define

$$\mathcal{F}(P_{XX'}) \triangleq \Big\{N(P_{XX'}) \leq e^{n(\Gamma(P_{XX'},R-\varepsilon)+R-E_0)}\Big\}. \tag{306}$$

Then, from (304) and (306), we obtain

$$\mathbb{P}[\mathcal{E}_0] \doteq \mathbb{P}\bigg[\bigcap_{P_{XX'}\in\mathcal{Q}(Q_X): d(P_{XX'})>\Delta}\mathcal{F}(P_{XX'})\bigg] \tag{307}$$

$$= \mathbb{P}\bigg[\bigcap_{P_{XX'}\in\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3}\mathcal{F}(P_{XX'})\bigg] \tag{308}$$

$$= \mathbb{P}\bigg[\bigcap_{P_{XX'}\in\mathcal{A}_3}\mathcal{F}(P_{XX'})\cap\bigcap_{P_{XX'}\in\mathcal{A}_1\cup\mathcal{A}_2}\mathcal{F}(P_{XX'})\bigg] \tag{309}$$

$$\geq \mathbb{P}\bigg[\bigcap_{P_{XX'}\in\mathcal{A}_3}\mathcal{F}(P_{XX'})\cap\mathcal{F}_0\bigg] \tag{310}$$

$$= \mathbb{P}\bigg[\bigcap_{P_{XX'}\in\mathcal{A}_3}\mathcal{F}(P_{XX'})\big|\mathcal{F}_0\bigg]\mathbb{P}[\mathcal{F}_0] \tag{311}$$

$$= \bigg(1-\mathbb{P}\bigg[\bigcup_{P_{XX'}\in\mathcal{A}_3}\mathcal{F}^c(P_{XX'})\big|\mathcal{F}_0\bigg]\bigg)\mathbb{P}[\mathcal{F}_0] \tag{312}$$

$$\geq \mathbb{P}[\mathcal{F}_0]-\sum_{P_{XX'}\in\mathcal{A}_3}\mathbb{P}\big[\mathcal{F}^c(P_{XX'})\big|\mathcal{F}_0\big]\mathbb{P}[\mathcal{F}_0] \tag{313}$$

$$\geq \mathbb{P}[\mathcal{F}_0]-\sum_{P_{XX'}\in\mathcal{A}_3}\mathbb{P}\big[\mathcal{F}^c(P_{XX'})\cap\mathcal{F}_0\big] \tag{314}$$

$$\geq \mathbb{P}[\mathcal{F}_0]-\sum_{P_{XX'}\in\mathcal{A}_3}\mathbb{P}\big[\mathcal{F}^c(P_{XX'})\big], \tag{315}$$

where (310) follows from the fact that for each joint type $P_{XX'} \in \mathcal{A}_1\cup\mathcal{A}_2$, it holds that $\{N(Q_{XX'})=0\}\subset\{N(Q_{XX'})\leq e^{n(\Gamma(P_{XX'},R-\varepsilon)+R-E_0)}\}$.

Equation (315) resembles [15, Eq. (205)] with subtle differences in the definition of sets $\mathcal{A}_1,\mathcal{A}_2$ and $\mathcal{A}_3$. However, since all the codewords in RGV are dependent, [15, Eq. (218)] does not hold. We proceed with different arguments. For any $P_{XX'}\in\mathcal{A}_3$, we have

$$\mathbb{P}\big[\mathcal{F}^c(P_{XX'})\big\} = \mathbb{P}\Big\{N(P_{XX'}) \geq e^{n(\Gamma(P_{XX'},R-\varepsilon)+R-E_0)}\Big] \tag{316}$$

$$\leq \mathbb{P}\Big[N(P_{XX'}) \geq e^{n\varepsilon}e^{n(2R-I_P(X;X'))}\Big], \tag{317}$$

where (317) follows from the definition of the set $\mathcal{A}_3$, which implies that

$$\Gamma(P_{XX'},R-\varepsilon)+R-E_0 > 2R-I_P(X;X')+\varepsilon. \tag{318}$$

On the other hand, by Lemma 7, we have

$$\sum_{P_{XX'}\in\mathcal{A}_3}\mathbb{P}\Big[N(P_{XX'}) \geq e^{n\varepsilon}e^{n(2R-I_P(X;X'))}\Big] \overset{\circ}{\leq} \max_{P_{XX'}\in\mathcal{A}_3}\exp\Big\{-e^{nR(2R-I_P(X;X')+\varepsilon)}\Big\} \tag{319}$$

$$= \exp\Big\{-e^{n(2R-\max_{P_{XX'}\in\mathcal{A}_3}I_P(X;X')+\varepsilon)}\Big\} \tag{320}$$

$$\leq \exp\Big\{-e^{n(2R-\min_{P_{XX'}\in\mathcal{A}_2}I_P(X;X')+\varepsilon)}\Big\}, \tag{321}$$

where (321) follows from the condition (267).

Now, under the condition (268), by Lemma 12, we have

$$\mathbb{P}\{F_0\} \overset{\circ}{\geq} \exp\big\{-e^{n\max_{P_{XX'}\in\mathcal{A}_2}(2R-I_P(X;X'))}\big\}. \tag{322}$$

From (315), (321), and (322), we obtain

$$\mathbb{P}[\mathcal{E}_0] \overset{\circ}{\geq} \exp\Big\{-e^{n\max_{P_{XX'}\in\mathcal{A}_2}(2R-I_P(X;X'))}\Big\} - \exp\Big\{-e^{n(2R-\min_{P_{XX'}\in\mathcal{A}_2}I_P(X;X')+\varepsilon)}\Big\} \tag{323}$$

$$\doteq \exp\Big\{-e^{n\max_{P_{XX'}\in\mathcal{A}_2}(2R-I_P(X;X'))}\Big\}. \tag{324}$$

To bound $\mathbb{P}[\mathcal{B}_{\varepsilon}(\tilde{m},\boldsymbol{y})\cap\mathcal{E}_0]$, we use the following arguments. As [15], let

$$\mathcal{N}^2 := \Big\{(m,m'): m\neq m', m,m'\in\{1,2,\cdots,\lfloor M/2\rfloor-1\}\Big\}. \tag{325}$$

Define

$$\mathcal{S} := \Big\{(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{\lfloor M\rfloor/2\rfloor})\in\underbrace{\mathbb{R}^n\times\mathbb{R}^n\cdots\times\mathbb{R}^n}_{\lfloor M/2\rfloor\ \text{times}}: \min_{i,j\in\{1,2,\cdots,\lfloor M/2\rfloor\},i\neq j}\{d(\boldsymbol{x}_i,\boldsymbol{x}_j)\}>\Delta\Big\}. \tag{326}$$

Since the distance between two codewords in a RGV ensemble is at least Δ, we have

$$\mathbb{P}[\mathcal{B}_{\varepsilon}(\tilde{m},\boldsymbol{y})\cap\mathcal{E}_0]$$

$$\leq \mathbb{P}\bigg[\sum_{(m,m')\in\mathcal{N}^2} e^{-n\Gamma(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}},R-\varepsilon)} \leq e^{n(R-E_0)}\bigg]$$
$$\times \mathbb{P}\bigg[\sum_{m'\in\{\lfloor M/2\rfloor,\cdots,M\}\setminus\{\tilde{m}\}} e^{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})} \leq e^{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})}$$
$$\bigg|\bigg\{\sum_{(m,m')\in\mathcal{N}^2} e^{-n\Gamma(\hat{P}_{\boldsymbol{X}_m,\boldsymbol{X}_{m'}},R-\varepsilon)} \leq e^{n(R-E_0)}\bigg\}$$
$$\cap\bigg\{(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_{\lfloor M\rfloor/2\rfloor})\in\mathcal{S}\bigg\}\bigg]. \tag{327}$$

Now, for any tuple $(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{\lfloor M\rfloor})$ such that $\min_{i,j\in\{1,2,\cdots,\lfloor M/2\rfloor\},i\neq j}\{d(\boldsymbol{x}_i,\boldsymbol{x}_j)\}>\Delta$, it holds that

$$\mathbb{P}\bigg(\boldsymbol{X}_{\lfloor M/2\rfloor+1}=\boldsymbol{x}_{\lfloor M/2\rfloor+1},\boldsymbol{X}_{\lfloor M/2\rfloor+2}=\boldsymbol{x}_{\lfloor M/2\rfloor+2},\cdots,$$
$$\boldsymbol{X}_M=\boldsymbol{x}_M\big|\boldsymbol{X}_1=\boldsymbol{x}_1,\cdots,\boldsymbol{X}_{\lfloor M/2\rfloor}=\boldsymbol{x}_{\lfloor M/2\rfloor}\bigg)$$
$$=\frac{\mathbb{P}(\boldsymbol{X}_1=\boldsymbol{x}_1,\boldsymbol{X}_2=\boldsymbol{x}_2,\cdots,\boldsymbol{X}_M=\boldsymbol{x}_M)}{\mathbb{P}(\boldsymbol{X}_{\lfloor M/2\rfloor}=\boldsymbol{x}_{\lfloor M/2\rfloor},\cdots,\boldsymbol{X}_1=\boldsymbol{x}_1)} \tag{328}$$
$$\dot{\leq}\frac{1}{|\mathcal{T}(Q_X)|^{\lceil M/2\rceil}}, \tag{329}$$

where (329) follows from Lemma 4. Hence, by using the same arguments as the proof of Lemma 14, we obtain

$$\mathbb{P}[\mathcal{B}_\varepsilon(\tilde{m},\boldsymbol{y})\cap\mathcal{E}_0]$$
$$\leq\exp\bigg\{-e^{n\varepsilon}\bigg[1-\frac{e^{-n(\varepsilon+\delta)}}{1-e^{-n\delta}}-e^{-n\varepsilon}(1+n\varepsilon)\bigg]\bigg\}\mathbb{P}(\mathcal{E}_0). \tag{330}$$

From (302), (324), and (330), we have

$$\mathbb{P}\bigg[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)\geq E_0\bigg]$$
$$\overset{\circ}{\geq}\bigg(1-e^{nR}|\mathcal{Y}|^n\exp\bigg\{-e^{n\varepsilon}\bigg[1-\frac{e^{-n(\varepsilon+\delta)}}{1-e^{-n\delta}}$$
$$-e^{-n\varepsilon}(1+n\varepsilon)\bigg]\bigg\}\bigg)$$
$$\times\exp\bigg\{-e^{n\max_{P_{XX'}\in\mathcal{A}_2}(2R-I_P(X;X'))}\bigg\} \tag{331}$$
$$\overset{\circ}{=}\exp\bigg\{-e^{n\max_{P_{XX'}\in\mathcal{A}_2}(2R-I_P(X;X'))}\bigg\} \tag{332}$$

which concludes the proof.

C. Convergence in Probability

This section enumerates properties of the tail exponents derived in Sections V-A and V-B, respectively, and establishes the convergence in probability to the TRC exponent of the RGV. In particular, the following results can be obtained by using the same arguments as the proofs of [15, Prop. 1], [15, Prop. 3], [15, Prop. 2], respectively, and are therefore stated without proof. Define

$$\tilde{E}(R)\triangleq\min_{P_{XX'}\in\mathcal{Q}(Q_X):I_P(X;X')\leq 2R,d(P_{XX'})>\Delta}\big\{\Lambda(P_{XX'},R)$$
$$+I_P(X;X')-R\big\}. \tag{333}$$

Proposition 3 (Lower Tail): $E_{\mathrm{lt}}^{\mathrm{ub}}(R,E_0)$ and $E_{\mathrm{lt}}^{\mathrm{lb}}(R,E_0)$ have the following properties

1) For fixed R, $E_{\mathrm{lt}}^{\mathrm{ub}}(R,E_0)$ and $E_{\mathrm{lt}}^{\mathrm{lb}}(R,E_0)$ are decreasing in E_0.
2) $E_{\mathrm{lt}}^{\mathrm{ub}}(R,E_0)>0$ if and only if $E_0<E_{\mathrm{trc}}^{\mathrm{rgv}}(R,Q_X,\Delta,d)$.
3) $E_{\mathrm{lt}}^{\mathrm{lb}}(R,E_0)>0$ if $E_0<\tilde{E}(R)$.
4) $E_{\mathrm{lt}}^{\mathrm{lb}}(R,E_0)=\infty$ for any $E_0<E_0^{\min}(R)$, where

$$E_0^{\min}(R)\triangleq\min_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})>\Delta}\big\{\Gamma(P_{XX'},R)$$
$$-[2R-I_P(X;X')]_++R\big\}. \tag{334}$$

Proposition 4 (Upper Tail): $E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0)$ and $E_{\mathrm{ut}}^{\mathrm{lb}}(R,E_0)$ have the following properties

1) For fixed R, $E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0)$ and $E_{\mathrm{ut}}^{\mathrm{lb}}(R,E_0)$ are increasing in E_0.
2) $E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0)>0$ if and only if $E_0>E_{\mathrm{trc}}^{\mathrm{rgv}}(R,Q_X,\Delta,d)$.
3) $E_{\mathrm{ut}}^{\mathrm{lb}}(R,E_0)>0$ if $E_0>\tilde{E}(R)$.

From Propositions 3 and 4, the following result states the convergence in probability to the TRC of the RGV ensemble.

Corollary 3: For any RGV ensemble with GLD, under the conditions in Lemma 11 and Lemma 12, we have that

$$-\frac{1}{n}\log P_{\mathrm{e}}(c_n)\xrightarrow{(\mathrm{p})}E_{\mathrm{trc}}^{\mathrm{rgv}}(R,Q_X,g,d,\Delta). \tag{335}$$

Recall that for $d(P_{XX'})=-I_P(X;X')$ and $\Delta=-(R+2\delta)$, the conditions in Lemma 11 and Lemma 12 hold. Hence, Corollary 3 holds for this important case for which $E_{\mathrm{rce}}^{\mathrm{rgv}}(R,Q_X,g,d,\Delta)=E_{\mathrm{trc}}^{\mathrm{rgv}}(R,Q_X,g,d,\Delta)=E_{\mathrm{ex}}^{\mathrm{cc}}(R,Q_X)$.

VI. Conclusion

We have studied the RGV code ensemble and have studied the typical error exponent and upper and lower error exponent tails. We have shown that the lower tail decays exponentially while the upper tail exhibits a decay that is between exponential and double-exponential; it is sub-double-exponential below the expurgated exponent and double-exponential above the expurgated exponent. In addition, we have shown that the error exponent of a sufficiently long RGV code concentrates in probability around the typical error exponent; this is also shown to coincide with the random coding exponent of the RGV ensemble, known to coincide with the maximum of the expurgated and the random-coding exponent. This suggests that every code in the ensemble asymptotically attains as high an error exponent as it is known from random codes.

Appendix A
Proof of Lemma 4

Assume that $\mathcal{A}=\{i_1,i_2,\cdots,i_l\}$ where $1\leq i_1<i_2<\cdots<i_l\leq M$ for some $l\in[M]$. First, if $\min_{j,k\in[l],j\neq k}d(\boldsymbol{x}_{i_j},\boldsymbol{x}_{i_k})\leq\Delta$, then by the RGV generation, we have

$$\mathbb{P}(\boldsymbol{x}_{i_1},\boldsymbol{x}_{i_2},\cdots,\boldsymbol{x}_{i_l})=0. \tag{336}$$

Hence, (36) trivially holds.

Now, under the condition $\min_{j,k\in[l],j\neq k} d(\boldsymbol{x}_{i_j},\boldsymbol{x}_{i_k}) > \Delta$, we have

$$\mathbb{P}\bigg[\bigcap_{k\in A}\{\boldsymbol{X}_k=\boldsymbol{x}_k\}\bigg] = \mathbb{P}(\boldsymbol{x}_{i_1},\boldsymbol{x}_{i_2},\cdots,\boldsymbol{x}_{i_l}) \tag{337}$$

$$= \sum_{x_1^{i_1-1},x_{i_1+1}^{i_2-1},\cdots,x_{i_{l-1}+1}^{i_l-1}:d(\boldsymbol{x}_k,\boldsymbol{x}_l)>\Delta\forall k,l\in[i_l],k\neq l} \mathbb{P}(\boldsymbol{x}_1^{i_1-1}) \times \mathbb{P}(\boldsymbol{x}_{i_1}|\boldsymbol{x}_1^{i_1-1})\mathbb{P}(\boldsymbol{x}_{i_1+1}^{i_2-1}|\boldsymbol{x}_1^{i_1})\mathbb{P}(\boldsymbol{x}_{i_2}|\boldsymbol{x}_1^{i_2-1}) \times \mathbb{P}(\boldsymbol{x}_{i_2+1}^{i_3-1}|\boldsymbol{x}_1^{i_2})\mathbb{P}(\boldsymbol{x}_{i_3}|\boldsymbol{x}_1^{i_3-1})\cdots \times \mathbb{P}(\boldsymbol{x}_{i_{l-1}+1}^{i_l-1}|\boldsymbol{x}_1^{i_{l-1}})\mathbb{P}(\boldsymbol{x}_{i_l}|\boldsymbol{x}_1^{i_l-1}) \tag{338}$$

$$= \sum_{x_1^{i_1-1},x_{i_1+1}^{i_2-1},\cdots,x_{i_{l-1}+1}^{i_l-1}:d(\boldsymbol{x}_k,\boldsymbol{x}_l)>\Delta\forall k,l\in[i_l],k\neq l} \mathbb{P}(\boldsymbol{x}_1^{i_1-1}) \times \mathbb{P}(\boldsymbol{x}_{i_1+1}^{i_2-1}|\boldsymbol{x}_1^{i_1})\mathbb{P}(\boldsymbol{x}_{i_2+1}^{i_3-1}|\boldsymbol{x}_1^{i_2})\cdots \times \mathbb{P}(\boldsymbol{x}_{i_{l-1}+1}^{i_l-1}|\boldsymbol{x}_1^{i_{l-1}})\prod_{j=1}^{l}\mathbb{P}(\boldsymbol{x}_{i_j}|\boldsymbol{x}_1^{i_j-1}) \tag{339}$$

$$= \sum_{x_1^{i_1-1},x_{i_1+1}^{i_2-1},\cdots,x_{i_{l-1}+1}^{i_l-1}:d(\boldsymbol{x}_k,\boldsymbol{x}_l)>\Delta\forall k,l\in[i_l],k\neq l} \mathbb{P}(\boldsymbol{x}_1^{i_1-1}) \times \mathbb{P}(\boldsymbol{x}_{i_1+1}^{i_2-1}|\boldsymbol{x}_1^{i_1})\mathbb{P}(\boldsymbol{x}_{i_2+1}^{i_3-1}|\boldsymbol{x}_1^{i_2})\cdots \times \mathbb{P}(\boldsymbol{x}_{i_{l-1}+1}^{i_l-1}|\boldsymbol{x}_1^{i_{l-1}})\prod_{j=1}^{l}\frac{1}{|\mathcal{T}(Q_X,\boldsymbol{x}_1^{i_j-1})|}. \tag{340}$$

On the other hand, under the condition 28, by Lemma 1, we have

$$|\mathcal{T}(Q_X)| \geq \mathcal{T}(Q_X,\boldsymbol{x}_1^{i-1})| \geq (1-e^{-n\delta})|\mathcal{T}(Q_X)|, \forall i\in[M] \tag{341}$$

for all $\boldsymbol{x}_1^{i-1}$ occurring with non-zero probability.

From (340) and (341), if $\min_{j,k\in[l],j\neq k} d(\boldsymbol{x}_{i_j},\boldsymbol{x}_{i_k}) > \Delta$, we obtain

$$\mathbb{P}\bigg[\bigcap_{k\in A}\{\boldsymbol{X}_k=\boldsymbol{x}_k\}\bigg] \leq \sum_{x_1^{i_1-1},x_{i_1+1}^{i_2-1},\cdots,x_{i_{l-1}+1}^{i_l-1}:d(\boldsymbol{x}_k,\boldsymbol{x}_l)>\Delta\forall k,l\in[i_l],k\neq l} \mathbb{P}(\boldsymbol{x}_1^{i_1-1}) \times \mathbb{P}(\boldsymbol{x}_{i_1+1}^{i_2-1}|\boldsymbol{x}_1^{i_1})\mathbb{P}(\boldsymbol{x}_{i_2+1}^{i_3-1}|\boldsymbol{x}_1^{i_2})\cdots \times \mathbb{P}(\boldsymbol{x}_{i_{l-1}+1}^{i_l-1}|\boldsymbol{x}_1^{i_{l-1}})\frac{1}{(1-e^{-n\delta})^l|\mathcal{T}(Q_X)|^l} \tag{342}$$

$$\leq \sum_{x_1^{i_1-1},x_{i_1+1}^{i_2-1},\cdots,x_{i_{l-1}+1}^{i_l-1}} \mathbb{P}(\boldsymbol{x}_1^{i_1-1})\mathbb{P}(\boldsymbol{x}_{i_1+1}^{i_2-1}|\boldsymbol{x}_1^{i_1})\mathbb{P}(\boldsymbol{x}_{i_2+1}^{i_3-1}|\boldsymbol{x}_1^{i_2})\cdots \times \mathbb{P}(\boldsymbol{x}_{i_{l-1}+1}^{i_l-1}|\boldsymbol{x}_1^{i_{l-1}})\frac{1}{(1-e^{-n\delta})^l|\mathcal{T}(Q_X)|^l} \tag{343}$$

$$= \frac{1}{(1-e^{-n\delta})^l|\mathcal{T}(Q_X)|^l} \tag{344}$$

$$= \frac{1}{(1-e^{-n\delta})^{|\mathcal{A}|}|\mathcal{T}(Q_X)|^{|\mathcal{A}|}}, \tag{345}$$

where (344) follows by summing over $\boldsymbol{x}_{i_{k-1}+1}^{i_k-1}$ for the k-th conditional distribution, and (345) follows from $|\mathcal{A}|=l$.

In addition, for any $M'\leq M$, from (340) and (341), if $\min_{k,l\in[M']:k\neq l} d(\boldsymbol{x}_k,\boldsymbol{x}_l) > \Delta$, we also have

$$\mathbb{P}\bigg[\bigcap_{k\in[M']}\{\boldsymbol{X}_k=\boldsymbol{x}_k\}\bigg] \geq \sum_{x_1^{i_1-1},x_{i_1+1}^{i_2-1},\cdots,x_{i_{l-1}+1}^{i_l-1}} \mathbb{P}(\boldsymbol{x}_1^{i_1-1})\mathbb{P}(\boldsymbol{x}_{i_1+1}^{i_2-1}|\boldsymbol{x}_1^{i_1})\mathbb{P}(\boldsymbol{x}_{i_2+1}^{i_3-1}|\boldsymbol{x}_1^{i_2}) \cdots \times \mathbb{P}(\boldsymbol{x}_{i_{l-1}+1}^{i_l-1}|\boldsymbol{x}_1^{i_{l-1}})\frac{1}{|\mathcal{T}(Q_X)|^{M'}} \tag{346}$$

$$= \frac{1}{|\mathcal{T}(Q_X)|^{M'}}. \tag{347}$$

This concludes our proof of Lemma 4.

Appendix B
Proof of Lemma 5

First, we prove (42). Observe that

$$\mathbb{E}[\mathcal{I}(i,j)] = \mathbb{P}\big[(\boldsymbol{X}_i,\boldsymbol{X}_j)\in\mathcal{T}(P_{XX'})\big] \tag{348}$$

$$= \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{T}(P_{XX'})}\mathbb{P}(\boldsymbol{x}_i,\boldsymbol{x}_j). \tag{349}$$

Now, let

$$\delta_n \triangleq \frac{e^{-n\delta}}{1-e^{-n\delta}}. \tag{350}$$

Then, under the condition (28) and $d(P_{XX'})>\Delta$, by Lemma 2, we have

$$\frac{(1-4\delta_n^2)}{|\mathcal{T}(Q_X)|^2}e^{-2\delta_n} \leq \mathbb{P}(\boldsymbol{x}_i,\boldsymbol{x}_j) \leq \frac{1}{(1-e^{-n\delta})^2|\mathcal{T}(Q_X)|^2} \tag{351}$$

for all $(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{T}(P_{XX'})$ since $d(\boldsymbol{x}_i,\boldsymbol{x}_j)=d(P_{XX'})>\Delta$. From (349) and (351), we have

$$(1-4\delta_n^2)e^{-2\delta_n}\frac{|\mathcal{T}(P_{XX'})|}{|\mathcal{T}(Q_X)|^2} \leq \mathbb{E}[\mathcal{I}(i,j)] \leq \frac{1}{(1-e^{-n\delta})^2}\frac{|\mathcal{T}(P_{XX'})|}{|\mathcal{T}(Q_X)|^2}. \tag{352}$$

Recall the definition of $L(P_{XX'})$ in (39). From (352), we have

$$(1-4\delta_n^2)e^{-2\delta_n}L(P_{XX'}) \leq \mathbb{E}[\mathcal{I}(i,j)] \leq \frac{1}{(1-e^{-n\delta})^2}L(P_{XX'}). \tag{353}$$

Now, we prove (43). We consider three cases:

- Case 1: $i=k, j\neq l$. Observe that

$$\mathbb{E}[\mathcal{I}(i,j)\mathcal{I}(i,l)] = \mathbb{P}\big[(\boldsymbol{X}_i,\boldsymbol{X}_j)\in\mathcal{T}(P_{XX'}),(\boldsymbol{X}_i,\boldsymbol{X}_l)\in\mathcal{T}(P_{XX'})\big] \tag{354}$$

$$= \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_l)\in\mathcal{T}^3(Q_X)}\mathbb{P}(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_l) \times \mathbb{1}\{\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{T}(P_{XX'})\}\cap\{(\boldsymbol{x}_i,\boldsymbol{x}_l)\in\mathcal{T}(P_{XX'})\}\} \tag{355}$$

$$\leq \frac{1}{(1-e^{-n\delta})^3}\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_l)\in\mathcal{T}^3(Q_X)}\mathbb{P}(\boldsymbol{x}_i)\mathbb{P}(\boldsymbol{x}_j)\mathbb{P}(\boldsymbol{x}_l)$$

$$\times \mathbb{1}\{\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{T}(P_{XX'})\}\cap\{(\boldsymbol{x}_i,\boldsymbol{x}_l)\in\mathcal{T}(P_{XX'})\}\} \tag{356}$$

$$= \frac{1}{(1-e^{-n\delta})^3}\sum_{\boldsymbol{x}_i\in\mathcal{T}(Q_X)}\mathbb{P}(\boldsymbol{x}_i) \times \mathbb{P}\big[(\boldsymbol{x}_i,\boldsymbol{X}_j)\in\mathcal{T}(P_{XX'})\big]\mathbb{P}\big[(\boldsymbol{x}_i,\boldsymbol{X}_l)\in\mathcal{T}(P_{XX'})\big] \tag{357}$$

$$= \frac{1}{(1-e^{-n\delta})^3}\sum_{\boldsymbol{x}_i\in\mathcal{T}(Q_X)}\mathbb{P}(\boldsymbol{x}_i)L^2(P_{XX'}) \tag{358}$$

$$= \frac{1}{(1-e^{-n\delta})^3}L^2(P_{XX'}), \tag{359}$$

where (356) follows from Lemma 4 and Lemma 3.

- $i\neq k, j=l$. The proof is similar to Case 1.
- $i\neq k, j\neq l$. Then, we have

$$\mathbb{E}[\mathcal{I}(i,j)\mathcal{I}(k,l)] = \mathbb{P}\big[(\boldsymbol{X}_i,\boldsymbol{X}_j)\in\mathcal{T}(P_{XX'}),(\boldsymbol{X}_k,\boldsymbol{X}_l)\in\mathcal{T}(P_{XX'})\big] \tag{360}$$

$$= \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_k,\boldsymbol{x}_l)\in\mathcal{T}^4(Q_X)}\mathbb{P}(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_k,\boldsymbol{x}_l) \times \mathbb{1}\{\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{T}(P_{XX'})\}\cap\{(\boldsymbol{x}_k,\boldsymbol{x}_l)\in\mathcal{T}(P_{XX'})\}\} \tag{361}$$

$$\le \frac{1}{(1-e^{-n\delta})^4}\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_k,\boldsymbol{x}_l)\in\mathcal{T}^4(Q_X)}\mathbb{P}(\boldsymbol{x}_i) \times\mathbb{P}(\boldsymbol{x}_j)\mathbb{P}(\boldsymbol{x}_k)\mathbb{P}(\boldsymbol{x}_l) \times \mathbb{1}\{\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{T}(P_{XX'})\}\cap\{(\boldsymbol{x}_k,\boldsymbol{x}_l)\in\mathcal{T}(P_{XX'})\}\} \tag{362}$$

$$= \frac{1}{(1-e^{-n\delta})^4}\mathbb{P}\big[(\boldsymbol{X}_i,\boldsymbol{X}_j)\in\mathcal{T}(P_{XX'})\big] \times \mathbb{P}\big[(\boldsymbol{X}_k,\boldsymbol{X}_l)\in\mathcal{T}(P_{XX'})\big] \tag{363}$$

$$= \frac{1}{(1-e^{-n\delta})^4}L^2(P_{XX'}), \tag{364}$$

where (362) follows from Lemma 4 and Lemma 3.

From (359) and (364), for any pairs $(i,j)\in[M]_*^2$ and $(k,l)\in[M]_*^2$ such that $(i,j)\neq(k,l)$, we have

$$\mathbb{E}[\mathcal{I}(i,j)\mathcal{I}(k,l)]\le\frac{1}{(1-e^{-n\delta})^4}L^2(P_{XX'}), \tag{365}$$

and we obtain (41).

Finally, by [28], it is easy to see that

$$L(P_{XX'})\doteq e^{-nI_P(X;X')}. \tag{366}$$

Hence, we obtain (42) and (43) from (40) and (41), respectively.

This concludes our proof of Lemma 5.

Appendix C
Proof of Lemma 6

Observe that

$$\mathbb{E}[N(P_{XX'})] = \mathbb{E}\Big[\sum_m\sum_{m'\neq m}\mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\}\Big] \tag{367}$$

$$= \sum_m\sum_{m'\neq m}\Bigg\{\sum_{\substack{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'}):\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta}}\mathbb{P}(\boldsymbol{x}_m,\boldsymbol{x}_{m'}) + \sum_{\substack{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'}):\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\le\Delta}}\mathbb{P}(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\Bigg\}. \tag{368}$$

On the other hand, by Lemma 2, under the condition (28), it holds that

$$\mathbb{P}(\boldsymbol{x}_m,\boldsymbol{x}_{m'})=0 \tag{369}$$

if $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\le\Delta$, and

$$\frac{1-4\delta_n^2}{|\mathcal{T}(Q_X)|^2}e^{-2\delta_n}\le\mathbb{P}(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\le\frac{1}{(1-e^{-n\delta})^2|\mathcal{T}(Q_X)|^2} \tag{370}$$

if $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta$.

From (368), (369) and (370), for any joint type $P_{XX'}$ such that $d(P_{XX'})>\Delta$, we obtain

$$\mathbb{E}[N(P_{XX'})] \ge \sum_m\sum_{m'\neq m}\sum_{\substack{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'}):\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta}}\mathbb{P}(\boldsymbol{x}_m,\boldsymbol{x}_{m'}) \tag{371}$$

$$\ge \sum_m\sum_{m'\neq m}\sum_{\substack{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'}):\\ d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta}}\frac{1-4\delta_n^2}{|\mathcal{T}(Q_X)|^2}e^{-2\delta_n} \tag{372}$$

$$= M(M-1)\sum_{\substack{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\\ d(P_{XX'})>\Delta}}\frac{1-4\delta_n^2}{|\mathcal{T}(Q_X)|^2}e^{-2\delta_n} \tag{373}$$

$$= M(M-1)|\mathcal{T}(P_{XX'})|\frac{1-4\delta_n^2}{|\mathcal{T}(Q_X)|^2}e^{-2\delta_n} \tag{374}$$

$$\ge (n+1)^{-3|\mathcal{X}|^2}(1-4\delta_n^2)e^{-2\delta_n}e^{n(2R-I_P(X;X'))}, \tag{375}$$

where (375) follows from [28].

Then, as n sufficiently large, we have

$$\mathbb{P}\big[\mathcal{E}(P_{XX'})\big] = \mathbb{P}\big[N(P_{XX'})<(1-4\delta_n^2)e^{-2\delta_n} \times\exp\{n[2R-I_P(X;X')-\varepsilon]\}\big] \tag{376}$$

$$\le \mathbb{P}\big[N(P_{XX'})<e^{-n\varepsilon/2}\mathbb{E}[N(P_{XX'})]\big] \tag{377}$$

$$= \mathbb{P}\left[\frac{N(P_{XX'})}{\mathbb{E}[N(P_{XX'})]}-1<-(1-e^{-n\varepsilon/2})\right] \tag{378}$$

$$\le \frac{\mathrm{Var}(N(P_{XX'}))}{(1-e^{-n\varepsilon/2})^2\big(\mathbb{E}[N(P_{XX'})]\big)^2}, \tag{379}$$

where (377) follows from (375), and (379) follows from Cauchy-Schwarz inequality.

Now, let

$$\mathcal{I}(m,m')\triangleq\mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\}, \tag{380}$$

and

$$L(P_{XX'})\triangleq\frac{|\mathcal{T}(P_{XX'})|}{|\mathcal{T}(Q_X)|^2}. \tag{381}$$

Then, it holds that [28],

$$L(P_{XX'})\ge(n+1)^{-3|\mathcal{X}|}e^{-nI_P(X;X')}. \tag{382}$$

Hence, as n sufficiently large, we have

$$\begin{aligned}
&M(M-1)L(P_{XX'}) \\
&\geq (n+1)^{-3|\mathcal{X}|}e^{n(2R-\frac{1}{n}-I_P(X;X'))} && (383)\\
&\geq (n+1)^{-3|\mathcal{X}|}e^{n(\varepsilon-\frac{1}{n})} && (384)\\
&\geq e^{n\varepsilon/2}, && (385)
\end{aligned}$$

where (385) follows from $\varepsilon >> (\log n)/\sqrt{n}$.

In addition, for any two fixed pairs (m,m') and $(\tilde{m},\hat{m})$ in $[M]_*^2$ such that $(m,m')\neq(\tilde{m},\hat{m})$, by Lemma 5, we have

$$\begin{aligned}
(1-4\delta_n^2)e^{-2\delta_n}L(P_{XX'}) &\leq \mathbb{E}[\mathcal{I}(m,m')] \\
&\leq \frac{1}{(1-e^{-n\delta})^2}L(P_{XX'}), && (386)
\end{aligned}$$

and

$$\mathbb{E}[\mathcal{I}(m,m')\mathcal{I}(\tilde{m},\hat{m})] \leq \frac{1}{(1-e^{-n\delta})^4}L^2(P_{XX'}). \quad (387)$$

It follows that

$$\begin{aligned}
&\mathrm{Var}(N(P_{XX'})) = \mathbb{E}[N^2(P_{XX'})] - \big(\mathbb{E}[N(P_{XX'})]\big)^2 && (388)\\
&= \sum_{m,m',\tilde{m},\hat{m}} \mathbb{E}[\mathcal{I}(m,m')\mathcal{I}(\tilde{m},\hat{m})] - \big(\mathbb{E}[N(P_{XX'})]\big)^2 && (389)\\
&= \sum_{m,m'} \mathbb{E}[\mathcal{I}(m,m')] \\
&\quad + \sum_{(m,m')\neq(\tilde{m},\hat{m})} \mathbb{E}\big[\mathcal{I}(m,m')\mathcal{I}(\tilde{m},\hat{m})\big] - \big(\mathbb{E}[N(P_{XX'})]\big)^2 && (390)\\
&\leq M(M-1)\frac{1}{(1-e^{-n\delta})^2}L(P_{XX'}) \\
&\quad + M(M-1)[M(M-1)-1]\frac{1}{(1-e^{-n\delta})^4}L^2(P_{XX'}) \\
&\quad - \big(\mathbb{E}[N(P_{XX'})]\big)^2. && (391)
\end{aligned}$$

Now, let

$$\begin{aligned}
V_n &= M(M-1)\frac{1}{(1-e^{-n\delta})^2}L(P_{XX'}) + M(M-1) \\
&\qquad \times [M(M-1)-1]\frac{1}{(1-e^{-n\delta})^4}L^2(P_{XX'}), && (392)\\
V_d &= \big((1-4\delta_n^2)e^{-2\delta_n}M(M-1)L(P_{XX'})\big)^2. && (393)
\end{aligned}$$

Then, from (374), (379), and (391), as n sufficiently large, we have

$$\begin{aligned}
&\mathbb{P}\big[\mathcal{E}(P_{XX'})\big] \\
&\leq \frac{\mathrm{Var}(N(P_{XX'}))}{(1-e^{-n\varepsilon/2})^2\big(\mathbb{E}[N(P_{XX'})]\big)^2} && (394)\\
&\leq \frac{1}{(1-e^{-n\varepsilon/2})^2}\left[\frac{V_n}{V_d}-1\right] && (395)\\
&\leq \frac{1}{(1-e^{-n\varepsilon/2})^2}\Bigg[\frac{e^{4\delta_n}}{\big(1-4\delta_n^2\big)^2\big(1-e^{-n\delta}\big)^2} \\
&\quad \times \left(\frac{1}{M(M-1)L(P_{XX'})}\right) + \frac{e^{4\delta_n}}{(1-4\delta_n^2)^2(1-e^{-n\delta})^4} - 1\Bigg] && (396)\\
&\leq \frac{1}{(1-e^{-n\varepsilon/2})^2}\Bigg[\frac{e^{4\delta_n}}{\big(1-4\delta_n^2\big)^2\big(1-e^{-n\delta}\big)^2}e^{-n\varepsilon/2} \\
&\quad + \frac{e^{4\delta_n}}{(1-4\delta_n^2)^2(1-e^{-n\delta})^4} - 1\Bigg], && (397)
\end{aligned}$$

where (397) follows from (385).

APPENDIX D
PROOF OF LEMMA 7

It is clear that (48) holds if $I_P(X;X')=0$ since the LHS of this inequality is equal to 0 for this case. Now, we consider the case $I_P(X;X')>0$. Then, we can choose $\delta(\varepsilon)$ such that $0<\delta(\varepsilon)<<$ such that $I_P(X;X')>\delta(\varepsilon)$. With an abuse of notation, we assume that $\delta(\varepsilon)=\varepsilon$.

Now, observe that

$$N(P_{XX'}) = \sum_{m=1}^{M}\sum_{m'\neq m}\mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\}. \quad (398)$$

By Lemma 5, we have

$$\mathbb{E}[\mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\}] \doteq e^{-nI_P(X;X')} \quad (399)$$

for all $(m,m')\in[M]_*^2$, which leads to

$$\begin{aligned}
p &\triangleq \frac{1}{M(M-1)}\mathbb{E}[N(P_{XX'})] && (400)\\
&\doteq e^{-nI_P(X;X')}. && (401)
\end{aligned}$$

By choosing $t=e^{n(2R-I_P(X;X')+\varepsilon)}+1$, then it is clear that

$$M(M-1)p \leq t-1 < M(M-1)-1 \quad (402)$$

as n sufficiently large if $I_P(X;X')>0$ and choose ε such that $0<\varepsilon<<$. Then, by applying Lemma 15, we obtain

$$\begin{aligned}
&\mathbb{P}\big[N(P_{XX'})\geq e^{n(2R-I_P(X;X')+\varepsilon)}\big] \\
&\overset{\circ}{\leq}\ \exp\big\{-M(M-1)D\big(e^{-n(I_P(X;X')-\varepsilon)}\|e^{-nI_P(X;X')}\big)\big\}. && (403)
\end{aligned}$$

Now, by using the fact that $D(a\|b)\geq a\big(\log\frac{a}{b}-1\big)$ [29], we have

$$\begin{aligned}
&D\big(e^{-n(I_P(X;X')-\varepsilon)}\|e^{-nI_P(X;X')}\big) \\
&\qquad \geq e^{-n(I_P(X;X')-\varepsilon)}\big(n\varepsilon-1\big). && (404)
\end{aligned}$$

From (403) and (404), we obtain (48). Finally, (49) is a straightforward consequence of (48). This concludes our proof of Lemma 7.

APPENDIX E
PROOF OF LEMMA 8

Similar to the proof of Lemma 8, by applying Lemma 15 with $t=e^{n\varepsilon}$, we finally have

$$\begin{aligned}
&\mathbb{P}\big[N(P_{XX'})\geq e^{n\varepsilon}\big] \\
&\quad \overset{\circ}{\leq}\ \exp\big\{-M(M-1)D\big(e^{n(\varepsilon-2R)}\|e^{-nI_P(X;X')}\big)\big\}. && (405)
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&D\big(e^{n(\varepsilon-2R)}\|e^{-nI_P(X;X')}\big) \\
&\quad \geq e^{n(\varepsilon-2R)}\big(n(\varepsilon-2R+I_P(X;X')-1\big). && (406)
\end{aligned}$$

From (405) and (406), we obtain (50) and (51).

APPENDIX F
PROOF OF LEMMA 9

Observe that

$$\mathbb{E}[N(P_{XX'})] = \sum_{m=1}^{M} \sum_{m' \neq m} \mathbb{E}[\mathcal{I}(m,m')] \tag{407}$$

$$\doteq e^{n(2R - I_P(X;X'))} \tag{408}$$

where (408) follows from Lemma 5. An upper bound in (52) simply follows from Markov's inequality and (408).

To show the lower bound, we use Suen's correlation inequality [15, A]. However, the dependency graph is now different from the one in [15, Proof of Lemma 6]. In this new dependency graph, each vertex (i,j) is connected to all other vertices or $M(M-1)-1$ vertices. Using the results of Lemma 5, we have

$$\Theta := \frac{1}{2} \sum_{(i,j)\in[M]_*^2} \sum_{(k,l)\in[M]_*^2,(k,l)\neq(i,j)} \mathbb{E}[\mathcal{I}(i,j)\mathcal{I}(k,l)] \tag{409}$$

$$\dot{\leq} \frac{1}{2} e^{2nR} e^{2nR} e^{-2nI_P(X;X')} \tag{410}$$

$$\doteq e^{n(4R - 2I_P(X;X'))}, \tag{411}$$

and

$$\Omega = \max_{(i,j)\in[M]_*^2} \sum_{(k,l)\in[M]_*^2,(k,l)\neq(i,j)} \mathbb{E}[\mathcal{I}(k,l)] \tag{412}$$

$$\doteq e^{2nR} e^{-nI_P(X;X')} \tag{413}$$

$$\doteq e^{n(2R - I_P(X;X'))}. \tag{414}$$

In addition, we have

$$\Delta = \mathbb{E}[N(P_{XX'})] \tag{415}$$

$$\doteq e^{n(2R - I_P(X;X'))}. \tag{416}$$

From (411), (414), and (416), we obtain

$$\frac{\Delta^2}{8\Theta} \dot{\geq} 1, \tag{417}$$

and

$$\frac{\Delta}{6\Omega} \doteq 1. \tag{418}$$

Now, by [15, Eq. (A.6)], we have

$$\mathbb{P}[N(P_{XX'}) = 0]$$

$$\leq \exp\left\{ -\min\left(\frac{\Delta^2}{8\Theta}, \frac{\Delta}{6\Omega}, \frac{\Delta}{2} \right) \right\} \tag{419}$$

$$\dot{\leq} \exp\left\{ -\min\left(1, 1, \frac{1}{2} e^{n(2R - I_P(X;X'))} \right) \right\} \tag{420}$$

$$= \exp\left\{ -\frac{1}{2} e^{n(2R - I_P(X;X'))} \right\}, \tag{421}$$

where (421) follows from the assumption $I_P(X;X') \geq 2R$.

From (421), by using the same arguments as [15, Proof of Lemma 6], we obtain

$$\mathbb{P}\big[N(P_{XX'}) \geq 1\big] \dot{\geq} \exp\{n(2R - I_P(X;X'))\}, \tag{422}$$

which is compatible with the upper bound, proving Lemma 9.

APPENDIX G
PROOF OF LEMMA 10

From Lemma 7 and the fact that $0 = e^{-n\infty}$, it holds that

$$\mathbb{P}[N(P_{XX'}) \geq e^{ns}] \doteq \exp(-n\infty) \tag{423}$$

if $s > [2R - I_P(X;X')]_+$.

Now, for $s < [2R - I_P(X;X')]_+$ and $2R \leq I_P(X;X')$, then $s \leq 0$. It follows that

$$\mathbb{P}[N(P_{XX'}) \geq e^{ns}] = \mathbb{P}[N(P_{XX'}) \geq 1] \tag{424}$$

$$\doteq \exp\big\{ n(2R - I_P(X;X')) \big\} \tag{425}$$

$$= \exp\big\{ -n[I_P(X;X') - 2R]_+ \big\}, \tag{426}$$

where (425) follows from Lemma 9.

On the other hand, for $s < [2R - I_P(X;X')]_+$ and $2R > I_P(X;X')$, then we have

$$\mathbb{P}[N(P_{XX'}) \geq e^{ns}] \leq 1 \tag{427}$$

$$= \exp\big\{ -n[I_P(X;X') - 2R]_+ \big\}. \tag{428}$$

In addition, for this case, there exists $\varepsilon > 0$ such that $2\varepsilon \leq \min\{2R - I_P(X;X'), [2R - I_P(X;X')]_+ - s\}$. Hence, by applying Lemma 6, we have

$$\mathbb{P}\big[N(P_{XX'}) \geq (1 - 4\delta_n^2) e^{-2\delta_n} \exp\{n[2R - I_P(X;X') - \varepsilon]\}\big] \to 1. \tag{429}$$

Furthermore, as n sufficiently large, we also have

$$\mathbb{P}[N(P_{XX'}) \geq e^{ns}]$$

$$\geq \mathbb{P}\big[N(P_{XX'}) \geq e^{n(2R - I_P(X;X') - 2\varepsilon)}\big] \tag{430}$$

$$\geq \mathbb{P}\big[N(P_{XX'}) \geq (1 - 4\delta_n^2) e^{-2\delta_n} \times \exp\{n[2R - I_P(X;X') - \varepsilon]\}\big] \tag{431}$$

$$= 1 + o(1) \tag{432}$$

$$= (1 + o(1)) \exp\big\{ -n[I_P(X;X') - 2R]_+ \big\} \tag{433}$$

$$\doteq \exp\big\{ -n[I_P(X;X') - 2R]_+ \big\}, \tag{434}$$

where (432) follows from (429), and (433) follows from $[I_P(X;X') - 2R]_+ = 0$ for $2R > I_P(X;X')$.

From (428) and (434), we obtain

$$\mathbb{P}[N(P_{XX'}) \geq e^{ns}] \doteq \exp\big\{ -n[I_P(X;X') - 2R]_+ \big\} \tag{435}$$

for $s < [2R - I_P(X;X')]_+$ and $2R > I_P(X;X')$.

By combining (426) and (435), we have

$$\mathbb{P}[N(P_{XX'}) \geq e^{ns}] \doteq \exp\big\{ -n[I_P(X;X') - 2R]_+ \big\} \tag{436}$$

for all $s < [2R - I_P(X;X')]_+$.

Finally, from (423) and (436), we obtain

$$E(R,P,s) = \begin{cases} [I_P(X;X') - 2R]_+, & [2R - I_P(X;X')]+ > s \\ +\infty, & [2R - I_P(X;X')]_+ < s. \end{cases} \tag{437}$$

This concludes our proof of Lemma 10.

APPENDIX H
PROOF OF LEMMA 11

First, we prove the following auxiliary lemma.

Lemma 17: For any $x \in [0, M^{-1}]$, the following holds:

$$1-(1-x)^M < 2e^{-Mx} \tag{438}$$

as M sufficiently large.

Proof: [Proof of Lemma 17] Let $g(x) \triangleq 1-(1-x)^M - 2e^{-Mx}$. This function has positive first-order derivative, hence $g(x)$ is increasing. Hence, for any $x \in [0, M^{-1}]$, we have

$$g(x) \le g(M^{-1}) \tag{439}$$

$$= 1-\left(1-\frac{1}{M}\right)^M - \frac{2}{e} \tag{440}$$

$$\to 1-\frac{3}{e} \quad \text{as} \qquad M\to\infty \tag{441}$$

$$< 0, \tag{442}$$

where (441) follows from $\left(1+\frac{1}{x}\right)^{-x} \to 1/e$ as $x\to\infty$. This concludes our proof of Lemma 17. ∎

Now, we return to prove Lemma H. Observe that

$$N(P_{XX'}) = \sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\}. \tag{443}$$

It follows that

$$\mathbb{E}[N(P_{XX'})] = \sum_{m=1}^M \sum_{m'\neq m} \mathbb{P}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} \tag{444}$$

$$\doteq e^{n(2R-I_P(X;X'))}, \tag{445}$$

where (445) follows from Lemma 5. Then, we have

$$\mathbb{P}\bigg\{N(P_{XX'}) \le e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\bigg\}$$
$$\dot{\le} \mathbb{P}\bigg\{\sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} \le e^{n(2R-I_P(X;X')-\varepsilon)}\bigg\}. \tag{446}$$

We consider two cases:

- The condition (55) holds.

On the space $\underbrace{\mathcal{X}^n\times\mathcal{X}^n\cdots\times\mathcal{X}^n}_{M \text{ terms}}$ define a probability measure $\mathbb{P}_\Pi$ such that

$$\mathbb{P}_\Pi(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M) = \prod_{m=1}^M \mathbb{P}[\boldsymbol{X}_m = \boldsymbol{x}_m] \tag{447}$$

for all $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M) \in \underbrace{\mathcal{X}^n\times\mathcal{X}^n\cdots\times\mathcal{X}^n}_{M \text{ terms}}$. Then, for this case, for any $P_{XX'}\in\mathcal{D}$, we have

$$\mathbb{P}\bigg\{\sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} \le e^{n(2R-I_P(X;X')-\varepsilon)}\bigg\}$$
$$= \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M} \mathbb{P}(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M) \times \mathbf{1}\bigg\{\sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})\} \le e^{n(2R-I_P(X;X')-\varepsilon)}\bigg\} \tag{448}$$
$$\le \frac{1}{(1-e^{-n\delta})^M} \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M} \mathbb{P}_\Pi(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M) \times \mathbf{1}\bigg\{\sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})\} \le e^{n(2R-I_P(X;X')-\varepsilon)}\bigg\} \tag{449}$$
$$= e^{-e^{nR}\log(1-e^{-n\delta})} \times \mathbb{P}_\Pi\bigg\{\sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} \le e^{n(2R-I_P(X;X')-\varepsilon)}\bigg\} \tag{450}$$
$$\overset{\circ}{\le} e^{-e^{nR}\log(1-e^{-n\delta})} \times \exp\bigg\{-\min\Big(e^{n(2R-I_P(X;X'))}, e^{nR}\Big)\bigg\}, \tag{451}$$

where (449) follows from Lemma 4, and (451) follows from [15, Lemma 2].

From (446) and (451), we obtain

$$\min_{P_{XX'}\in D} \mathbb{P}\bigg\{N(P_{XX'}) \le e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\bigg\}$$
$$\overset{\circ}{\le} e^{-e^{nR}\log(1-e^{-n\delta})} \times \exp\bigg\{-\min\Big(e^{n(2R-\min_{P_{XX'}\in D} I_P(X;X'))}, e^{nR}\Big)\bigg\} \tag{452}$$
$$\overset{\circ}{\le} e^{-e^{nR}\log(1-e^{-n\delta})} \exp\big\{-e^{n(R-2\delta)}\big\} \tag{453}$$
$$\overset{\circ}{=} \exp\big\{-e^{n(R-2\delta)}\big\}, \tag{454}$$

where (453) follows from $\min_{P_{XX'}\in\mathcal{D}} I_P(X;X') \le R+2\delta$ for this case, and (454) follows from $-\log(1-e^{-n\delta}) \sim e^{-n\delta}$.

- Case 2: The condition (56) holds.

For this case, observe that

$$\mathbb{P}\bigg\{N(P_{XX'}) > e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\bigg\}$$
$$\ge \mathbb{P}\bigg\{\bigg\{\sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} > e^{n(2R-I_P(X;X')-\varepsilon)}\bigg\} \cap \bigg\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) > \Delta\bigg\}\bigg\} \tag{455}$$
$$= \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M} \mathbb{P}(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M) \times \mathbb{1}\bigg\{\bigg\{\sum_{m=1}^M \sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})\}$$

$$> e^{n(2R-I_P(X;X')-\varepsilon)}\Big\} \cap \Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\Big\}\Big\} \tag{456}$$

$$\geq \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M} \mathbb{P}_\Pi(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) \times \mathbb{1}\Big\{\Big\{\sum_{m=1}^M\sum_{m'\neq m}\mathbb{1}\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\} > e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\} \cap \Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\Big\}\Big\} \tag{457}$$

$$= \mathbb{P}_\Pi\Big\{\Big\{\sum_{m=1}^M\sum_{m'\neq m}\mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\} > e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\} \cap \Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta\Big\}\Big\}, \tag{458}$$

where (457) follows from Lemma 4 with $M' = M$ and Lemma 3.

From (458), we have

$$\mathbb{P}\Big\{N(P_{XX'}) \leq e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\}$$

$$\leq \Pr_\Pi\Big\{\Big\{\sum_{m=1}^M\sum_{m'\neq m}\mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\} \leq e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\} \cup \Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\leq\Delta\Big\}\Big\} \tag{459}$$

$$= \mathbb{P}_\Pi\Big\{\sum_{m=1}^M\sum_{m'\neq m}\mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\} \leq e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\} + \mathbb{P}_\Pi\Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\leq\Delta\Big\}. \tag{460}$$

Now, observe that

$$\Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\leq\Delta\Big\} = \Big\{\bigcup_{m=1}^M\bigcup_{m'\neq m}\{d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\leq\Delta\}\Big\} \tag{461}$$

$$= \Big\{\bigcup_{m=1}^M\bigcup_{m'\neq m}\bigcup_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(\tilde{P}_{XX'})\}\Big\}. \tag{462}$$

Therefore, we have

$$\mathbb{P}_\Pi\Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\leq\Delta\Big\} = \mathbb{P}_\Pi\Big\{\bigcup_{m=1}^M\bigcup_{m'\neq m}\bigcup_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(\tilde{P}_{XX'})\}\Big\} \tag{463}$$

$$\leq \sum_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta}\sum_{m=1}^M \mathbb{P}_\Pi\Big\{\bigcup_{m'\neq m}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(\tilde{P}_{XX'})\}\Big\}. \tag{464}$$

Now, for any joint-type $\tilde{P}_{XX'}\in\mathcal{Q}(Q_X)$ such that $d(\tilde{P}_{XX'})\leq\Delta$, we have

$$\mathbb{P}_\Pi\Big\{\bigcup_{m'\neq m}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(\tilde{P}_{XX'})\}\Big\}$$

$$= \mathbb{E}\Big[\mathbb{P}_\Pi\Big\{\bigcup_{m'\neq m}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(\tilde{P}_{XX'})\}\Big|\boldsymbol{X}_m\Big\}\Big] \tag{465}$$

$$= 1-\mathbb{E}\Big[\mathbb{P}_\Pi\Big\{\bigcap_{m'\neq m}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\notin\mathcal{T}(\tilde{P}_{XX'})\}\Big|\boldsymbol{X}_m\Big\}\Big] \tag{466}$$

$$= 1-\mathbb{E}\Big[\Big(\mathbb{P}_\Pi\{(\boldsymbol{X}_m,\boldsymbol{X}_{m \bmod M+1}) \notin\mathcal{T}(\tilde{P}_{XX'})\Big|\boldsymbol{X}_m\}\Big)^M\Big] \tag{467}$$

$$\doteq 1-\big(1-e^{-nI_{\tilde{P}}(X;X')}\big)^M, \tag{468}$$

where (468) follows from the standard calculation (eg. [28]).

Now, from the condition (56), we have

$$R \leq \min_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta} I_{\tilde{P}}(X;X')-2\delta, \tag{469}$$

which leads to

$$e^{-n\min_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta} I_{\tilde{P}}(X;X')} \leq e^{-nR} = M^{-1}. \tag{470}$$

From (464) and (468), we obtain

$$\mathbb{P}_\Pi\Big\{\min_{(m,m')\in[M]_*^2} d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\leq\Delta\Big\}$$

$$\dot{\leq} M\Big[1-\big(1-e^{-n\min_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta} I_{\tilde{P}}(X;X')}\big)^M\Big] \tag{471}$$

$$\mathring{\leq}\ 2M\exp\Big\{-Me^{-n\min_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta} I_{\tilde{P}}(X;X')}\Big\} \tag{472}$$

$$\mathring{\leq}\ \exp\Big\{-e^{n\big(R-\min_{\tilde{P}_{XX'}\in\mathcal{Q}(Q_X):d(\tilde{P}_{XX'})\leq\Delta} I_{\tilde{P}}(X;X')\big)}\Big\} \tag{473}$$

$$\mathring{\leq}\ \exp\Big\{-e^{n\big(2R+2\delta-\min_{\tilde{P}_{XX'}\in D} I_{\tilde{P}}(X;X')\big)}\Big\} \tag{474}$$

where (472) follows from Lemma 17 with (470), (474) follows from the condition (56).

On the other hand, by [15, Prep. 6], we have

$$\mathbb{P}_\Pi\Big\{N(P_{XX'}) \le e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\} \tag{475}$$

$$\doteq \mathbb{P}_\Pi\Big\{N(P_{XX'}) \le e^{-n\varepsilon}e^{n(2R-I_P(X;X'))}\Big\} \tag{476}$$

$$\overset{\circ}{\le} \exp\Big\{-e^{n(2R-I_P(X;X'))}\Big\}. \tag{477}$$

From (474) and (477), under the condition (56), we have

$$\min_{P_{XX'}\in D}\mathbb{P}\Big\{N(P_{XX'}) \le e^{-n\varepsilon}\mathbb{E}[N(P_{XX'})]\Big\}$$
$$\overset{\circ}{\le} \exp\Big\{-e^{n(2R-\min_{P_{XX'}\in D} I_P(X;X'))}\Big\}. \tag{478}$$

Finally, we obtain by combining (454) for the case 1 and (478) for the case 2.

This concludes our proof of Lemma H.

Appendix I
Proof of Lemma 12

Define a new probability measure Π on $\underbrace{\mathcal{X}^n \times \mathcal{X}^n \cdots \times \mathcal{X}^n}_{M \text{ times}}$:

$$\mathbb{P}_\Pi(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) = \prod_{m=1}^M \mathbb{P}[\boldsymbol{X}_m = \boldsymbol{x}_m] \tag{479}$$

for all $(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M)$.

Observe that

$$\mathbb{P}(F_0)$$
$$= \mathbb{P}\Big\{\sum_{P_{XX'}\in\mathcal{A}_1\cup\mathcal{A}_2} N(P_{XX'}) = 0\Big\} \tag{480}$$

$$= \mathbb{P}\Big\{\sum_{P_{XX'}\in\mathcal{A}_1\cup\mathcal{A}_2}\sum_{m=1}^M\sum_{m'\neq m} \mathbb{1}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\} = 0\Big\} \tag{481}$$

$$= \mathbb{P}\Big\{\bigcap_{P_{XX'}\in\mathcal{A}_1\cup\mathcal{A}_2}\bigcap_{m=1}^M\bigcap_{m'\neq m}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\}^c\Big\} \tag{482}$$

$$= \mathbb{P}\Big\{\bigcap_{P_{XX'}\in A_1\cup A_2}\bigcap_{m=1}^M\bigcap_{m'\neq m}\{\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\} \cap\{d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta\}\}^c\Big\} \tag{483}$$

$$= \mathbb{P}\Big\{\bigcap_{P_{XX'}\in A_1\cup A_2}\bigcap_{m=1}^M\bigcap_{m'\neq m}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\notin\mathcal{T}(P_{XX'})\} \cup\{d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\le\Delta\}\Big\} \tag{484}$$

$$= \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M}\mathbb{P}(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) \times\prod_{P_{XX'}\in A_1\cup A_2}\prod_{m=1}^M\prod_{m'\neq m}\mathbb{1}\Big\{\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\notin\mathcal{T}(P_{XX'})\} \cup\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\le\Delta\}\Big\} \tag{485}$$

$$= \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M}\mathbb{P}(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) \times\prod_{P_{XX'}\in A_1\cup A_2}\prod_{m=1}^M\prod_{m'\neq m}\Big(1-\mathbb{1}\big\{\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\} \cap\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\}\big\}\Big) \tag{486}$$

$$\ge \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M}\mathbb{P}(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) \times\prod_{P_{XX'}\in A_1\cup A_2}\prod_{m=1}^M\prod_{m'\neq m}\Big(1-\mathbb{1}\big\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\big\}\Big) \times\mathbb{1}\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\} \tag{487}$$

$$= \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M}\mathbb{P}(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) \times\prod_{P_{XX'}\in A_1\cup A_2}\prod_{m=1}^M\prod_{m'\neq m} \times\mathbb{1}\big\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\notin\mathcal{T}(P_{XX'})\big\}\mathbb{1}\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\} \tag{488}$$

$$\ge \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M}\mathbb{P}_\Pi(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) \times\Pi_{P_{XX'}\in A_1\cup A_2}\Pi_{m=1}^M\Pi_{m'\neq m} \times\mathbb{1}\big\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\notin\mathcal{T}(P_{XX'})\big\}\mathbb{1}\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\} \tag{489}$$

$$= \mathbb{P}_\Pi\Big\{\bigcap_{P_{XX'}\in A_1\cup A_2}\bigcap_{m=1}^M\bigcap_{m'\neq m}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\notin\mathcal{T}(P_{XX'})\} \cap\{d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})>\Delta\}\Big\} \tag{490}$$

$$= \mathbb{P}_\Pi\Big\{\sum_{P_{XX'}\in A_1\cup A_2}\sum_{m=1}^M\sum_{m'\neq m} \mathbb{1}\{\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\} \cup\{d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\le\Delta\}\} = 0\Big\}, \tag{491}$$

where (483) follows from $d(P_{XX'}) > \Delta$ for all $P_{XX'} \in A_1\cup A_2$ and $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'}) = d(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}})$, (487) follows from the fact that $1-\mathbb{1}\big\{\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\}\cap\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\}\big\} = \big(1-\mathbb{1}\big\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\big\}\big)\mathbb{1}\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\}$ if $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta$ and $1-\mathbb{1}\big\{\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\}\cap\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\}\big\} \ge 0 = \big(1-\mathbb{1}\big\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(P_{XX'})\big\}\big)\mathbb{1}\{d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})>\Delta\}$ if $d(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\le\Delta$, (489) follows from [7, Lemma 4] and Lemma 4.

To apply Lemma (21), we form a dependency graph as follows. Define the family of Bernoulli random variables $\{\mathcal{I}(m,m',P_{XX'})\}_{P_{XX'}\in\mathcal{A}_1\cup\mathcal{A}_2,(m,m')\in[M]^2_*}$, where

$$\mathcal{I}(m,m',P_{XX'}) \triangleq \mathbb{1}\big\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(P_{XX'})\cup\{d(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\le\Delta\}\big\}. \tag{492}$$

Then, we have

$$\mathbb{E}_\Pi[\mathcal{I}(m,m',P_{XX'})]$$

$$= \mathbb{P}_\Pi\big\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'}) \cup \{d(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \le \Delta\}\big\} \tag{493}$$

$$\le \mathbb{P}_\Pi\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} + \mathbb{P}_\Pi\big\{d(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \le \Delta\big\}. \tag{494}$$

On the other hand, we have

$$\begin{aligned}
&\mathbb{P}_\Pi\big\{d(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \le \Delta\big\} \\
&= \sum_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}} \mathbb{P}_\Pi(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \mathbb{1}\{d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \le \Delta\} && (495)\\
&= \sum_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}} \mathbb{P}(\boldsymbol{x}_m)\mathbb{P}(\boldsymbol{x}_{m'}) \mathbb{1}\{d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \le \Delta\} && (496)\\
&= \sum_{P_{XX'} \in \mathcal{Q}(Q_X)} \sum_{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})} \mathbb{P}(\boldsymbol{x}_m)\mathbb{P}(\boldsymbol{x}_{m'}) \\
&\qquad \times \mathbb{1}\{d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \le \Delta\} && (497)\\
&= \sum_{P_{XX'} \in \mathcal{Q}(Q_X)} \sum_{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})} \frac{1}{|T(Q_X)|^2} \\
&\qquad \times \mathbb{1}\{d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \le \Delta\} && (498)\\
&= \sum_{P_{XX'} \in \mathcal{Q}(Q_X)} \sum_{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(P_{XX'})} \frac{1}{|T(Q_X)|^2} \\
&\qquad \times \mathbb{1}\{d(P_{XX'}) \le \Delta\} && (499)\\
&\doteq \max_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \le \Delta} e^{-nI_P(X;X')} && (500)\\
&= e^{-n \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \le \Delta} I_P(X;X')} && (501)\\
&\le e^{-n \max_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) > \Delta} I_P(X;X')}, && (502)
\end{aligned}$$

where (498) follows from 3, and (502) holds by the condition (69) under (68).

It follows from (494) and (502) that

$$\begin{aligned}
&\mathbb{E}_\Pi[\mathcal{I}(m, m', P_{XX'})] \\
&\le \mathbb{P}_\Pi\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} \\
&\quad + e^{-n \max_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) > \Delta} I_P(X;X')} && (503)\\
&\doteq e^{-nI_P(X;X')} + e^{-n \max_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) > \Delta} I_P(X;X')} && (504)\\
&\le e^{-nI_P(X;X')} + e^{-nI_P(X;X')} && (505)\\
&\doteq e^{-nI_P(X;X')}, && (506)
\end{aligned}$$

where (505) follows from the fact that $d(P_{XX'}) > \Delta$ for all $P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2$.

Now, we set

$$x(m, m', P_{XX'}) \triangleq 1 - \exp\big\{-e^{nI_P(X;X')}\big\}. \tag{507}$$

Then, under the condition $\min_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} I_P(X;X') > R$, for all $(m, m', P_{XX'}) \in [M]_*^2 \times (\mathcal{A}_1 \cup \mathcal{A}_2)$, it holds that

$$\begin{aligned}
&\mathbb{E}_\Pi[\mathcal{I}(m, m', P_{XX'})] && (508)\\
&\dot{\le} e^{-nI_P(X;X')} && (509)\\
&\doteq 1 - \exp\Big\{-e^{-nI_P(X;X')}\Big\} && (510)\\
&\doteq \bigg(1 - \exp\Big\{-e^{-nI_P(X;X')}\Big\}\bigg) \\
&\quad \times \bigg(\exp\Big\{-e^{-nI_P(X;X')}\Big\}\bigg)^{|\mathcal{A}_1 \cup \mathcal{A}_2| e^{nR}} && (511)\\
&= x(m, m', P_{XX'}) \\
&\quad \times \prod_{(\tilde{m}, \tilde{m}', \tilde{P}_{XX'}) \sim (m, m', P_{XX'})} \Big(1 - x(\tilde{m}, \tilde{m}', \tilde{P}_{XX'})\Big), && (512)
\end{aligned}$$

where (510) follows from the fact that $\lim_{x \to 0} \frac{e^{-x}}{1-x} = 1$, (511) follows from $|\mathcal{A}_1 \cup \mathcal{A}_2| \le |\mathcal{Q}(Q_X)|$ which is sub-exponential in n and $\min_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} I_P(X;X') > R$.

Then, by applying Lemma 21 with $A = [M]_*^2 \times (\mathcal{A}_1 \cup \mathcal{A}_2)$ and $B = \emptyset$, under the condition $\min_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} I_P(X;X') > R$ we have

$$\begin{aligned}
&\mathbb{P}_\Pi\bigg\{\sum_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \sum_{m=1}^{M} \sum_{m' \neq m} \\
&\quad \mathbb{1}\big\{\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(P_{XX'})\} \\
&\qquad \cup \{d(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \le \Delta\}\big\} = 0\bigg\} \\
&\ge \min_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \bigg(\exp\Big\{-e^{nI_P(X;X')}\Big\}\bigg)^{|\mathcal{A}_1 \cup \mathcal{A}_2| M(M-1)} && (513)\\
&\overset{\circ}{\doteq} \exp\Big\{-e^{n \max_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2}(2R - I_P(X;X'))}\Big\} && (514)\\
&= \exp\Big\{-e^{n \max_{P_{XX'} \in \mathcal{A}_2}(2R - I_P(X;X'))}\Big\}, && (515)
\end{aligned}$$

where (515) follows from the definition of $\mathcal{A}_1$ and $\mathcal{A}_2$.

Finally, the condition $\min_{P_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} I_P(X;X') > R$ is the same as $\min_{P_{XX'} \in \mathcal{A}_2} I_P(X;X') > R$, which is equivalent to the condition that

$$\begin{aligned}
&E_0 < E_{\text{ex}}^g(R, Q_X, d, \Delta) \\
&\triangleq \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) > \Delta, I_P(X;X') \le R} \Big\{\Gamma(P_{XX'}, R) \\
&\qquad + I_P(X;X') - R\Big\} && (516)\\
&= E_{\text{ex}}^{\text{rgv}}(R, Q_X, g, d, \Delta), && (517)
\end{aligned}$$

where (516) is obtained by using the same arguments to achieve [15, I. (30)]. This concludes our proof of Lemma 12.

Appendix J
Concentration Inequalities for Sums of Bernoulli Random Variables

To obtain the TRC or develop concentration inequalities for the random coding exponents, we need to develop concentration inequalities for a sum of Bernoulli random variables. Since in RGV codebooks, all the codewords are correlated, standard concentration inequalities such as Suen's correlation inequality [15], [32] cannot be applied. The main reason is that these standard inequality require a local dependency in the sum of random variables which only holds for the fixed-composition or i.i.d. random ensembles but not for RGV ones. We develop concentration inequalities for a sum of n

terms where each term depends on all the $n-1$ other terms. Thanks to the structure of all these random variables, some concentration inequalities in the probability literature can be applied. In this section, we list all these inequalities. For the newly-developed inequality, the proof can be found in appendices.

Lemma 18: [31, Lemma 2.1] Fix a positive number n and let $\{x_1, x_2, \cdots, x_n\}$ be real numbers from the interval $[0,1]$. For every $A \subset [n]$, let ζ_A be defined as

$$\zeta_A = \prod_{i\in A} x_i \prod_{i\in[n]\setminus A} (1-x_i). \tag{518}$$

Then,

$$\sum_{A\subset[n]} \zeta_A = \sum_{j=0}^{n} \sum_{A\in\partial_j[n]} \zeta_A = 1 \tag{519}$$

and

$$\sum_{i=1}^{n} x_i = \sum_{j=0}^{n} j \sum_{A\in\partial_j[n]} \zeta_A, \tag{520}$$

where $\partial_j[n]$ denotes the family consisting of all subsets of $[n]$ of cardinality $j \in \{0,1,2,\cdots,n\}$.

The following result can be also derived from Lemma 15.

Lemma 19: Suppose that $X_1, X_2, \cdots, X_n$ are random variables such that $X_i \in \{0,1\}$, for $i = 1,2,\cdots,n$. Set $p = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]$. Then, for any $\nu \in [0,p)$, it holds that

$$\mathbb{P}\bigg[\sum_{i=1}^{n} X_i \le n(p-\nu)-1\bigg] \le 2\ e^{-nD(p-\nu\|p)}. \tag{521}$$

Proof: Let $\tilde{X}_i \triangleq 1 - X_i$ for all $i\in[n]$ and set $\tilde{p} \triangleq 1-p$. Then, we have

$$\tilde{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\tilde{X}_i]. \tag{522}$$

Let $t-1 = n(1-p) + n(1-p)\varepsilon_0$ for some $\varepsilon_0 > 0$ such that $(1-p)(1+\varepsilon_0) < 1$. Then, by applying Lemma 15 for the Bernoulli sequence $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n$, we have

$$\mathbb{P}\bigg[\sum_{i=1}^{n} \tilde{X}_i \ge t\bigg] \le 2\ e^{-nD(\tilde{p}(1+\varepsilon_0)\|\tilde{p})} \tag{523}$$

$$= 2e^{-nD((1-p)(1+\varepsilon_0)\|1-p)}. \tag{524}$$

From (524) and $\tilde{X}_i = 1 - X_i$ for all $i \in [n]$, we obtain

$$\mathbb{P}\bigg[\sum_{i=1}^{n} X_i \le n-t\bigg] \le 2\ e^{-nD((1-p)(1+\varepsilon_0)\|1-p)}. \tag{525}$$

Now, by setting $\varepsilon_0 \triangleq \nu/(1-p)$, we have $t = n(1-p+\nu)+1$. Then, from (525), we have

$$\mathbb{P}\bigg[\sum_{i=1}^{n} X_i \le n(p-\nu)-1\bigg] \le 2\ e^{-nD((1-p)(1+\varepsilon_0)\|1-p)} \tag{526}$$

$$= 2e^{-nD(1-p+\nu\|1-p)} \tag{527}$$

$$= 2e^{-nD(p-\nu\|p)}, \tag{528}$$

where (528) follows from $D(a\|b) = D(1-a\|1-b)$. Final note is that $(1-p)(1+\varepsilon_0) = 1-p+\nu < 1$ for all $\nu \in [0,p)$. ■

Now, we recall the following result.

Lemma 20 [31, Theorem 1.2]: There exists a universal constant $c \ge 1$ satisfying the following. Suppose $X_1, X_2, \cdots, X_n$ are random variables such that $0 \le X_i \le 1$, for $i = 1,2,\cdots,n$. Assume further that there exists constant $\gamma \in (0,1)$ such that for all $A \subset [n]$ the following condition holds true:

$$\mathbb{E}\bigg[\prod_{i\in A} X_i\bigg] \le \gamma^{|A|} \tag{529}$$

where $|A|$ denotes the cardinality of A. Fix a real number ν from the interval $\big(0, \frac{1}{\gamma}-1\big)$ and set $t = n\gamma + n\gamma\nu$. Then,

$$\mathbb{P}\bigg[\sum_{i=1}^{n} X_i \ge t\bigg] \le c e^{-nD(\gamma(1+\nu)\|\gamma)}, \tag{530}$$

where $D(\gamma(1+\nu)\|\gamma)$ is the Kullback-Leibler distance between $\gamma(1+\nu)$ and γ.

Now, to bound the probability in (491), we recall the following version of Suen's correlation inequality lemma in [32].

Lemma 21 [32, Lemma 1]: Let $\{U_k\}_{k\in\mathcal{K}}$, where $\mathcal{K}$ is a set of multidimensional indexes, be a family of Bernoulli random variables. Let G be a dependency graph for $\{U_k\}_{k\in\mathcal{K}}$, i.e., a graph with vertex set $\mathcal{K}$ such that if A and B are two disjoint subsets of $\mathcal{K}$, and G contains no edge between A and B, then the families $\{U_k\}_{k\in A}$ and $\{U_k\}_{k\in B}$ are independent. Let $S_A \triangleq \sum_{k\in A} U_k$ for any $A \subset \mathcal{K}$. Moreover, we write $k \sim l$ if (k,l) is an edge in the dependency graph G. Suppose further that $x_k, k \in \mathcal{K}$ are real numbers such that $0 \le x_k < 1$ and

$$\mathbb{E}[U_k] \le x_k \prod_{l\sim k} (1-x_l), \qquad k \in \mathcal{K}. \tag{531}$$

Then, for any two subsets $A, B \subset \mathcal{K}$, it holds that

$$\mathbb{P}\big(S_A = 0 | S_B = 0\big) \ge \prod_{i\in A} (1-x_i). \tag{532}$$

APPENDIX K
PROOF OF LEMMA 14

Fix an $m \in [M]$. For any conditional type $P_{X'Y} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$ such that $P_{X'} = Q_X$ and $P_Y = \hat{P}_{\boldsymbol{y}}$, define

$$N_{m,\boldsymbol{y}}(P_{X'Y}) \triangleq \big|\big\{\boldsymbol{X}_{m'} : (\boldsymbol{X}_{m'}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y}), m' \neq m\big\}\big| \tag{533}$$

$$= \sum_{m'\neq m} \mathbb{1}\big\{(\boldsymbol{X}_{m'}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y})\big\}. \tag{534}$$

Observe that

$$\mathbb{E}\big[\mathbb{1}\big\{(\boldsymbol{X}_{m'}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y})\big\}\big]$$

$$= \mathbb{P}\big[(\boldsymbol{X}_{m'}, \boldsymbol{y}) \in \mathcal{T}(P_{X'Y})\big] \tag{535}$$

$$= \sum_{\boldsymbol{x}'_m \in \mathcal{T}(P_{X'|Y})} \mathbb{P}(\boldsymbol{X}_{m'} = \boldsymbol{x}_{m'}) \tag{536}$$

$$= \sum_{\boldsymbol{x}'_m \in \mathcal{T}(P_{X'|Y})} \frac{1}{|\mathcal{T}(Q_X)|} \tag{537}$$

$$\doteq e^{-nI_P(X';Y)}, \tag{538}$$

where (537) follows from Lemma 3, and (538) follows from [28]. Hence, $N_{m,\boldsymbol{y}}(P_{X'Y})$ is a sum of $M-1$ binary-valued random variables, each has the expectation $e^{-nI(X';Y)}$.

Now, from (98) and (534), we can express $Z_m(\boldsymbol{y})$ as

$$Z_m(\boldsymbol{y}) = \sum_{P_{X'|Y}:P_{X'}=Q_X} N_{m,y}(P_{X'Y})e^{ng(P_{X'Y})}. \tag{539}$$

Hence, by considering the randomness of $\{\boldsymbol{X}_{m'}\}$, we have

$$\begin{aligned}
&\mathbb{P}\big[Z_m(\boldsymbol{y}) \le \exp\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\}\big] \\
&\le \mathbb{P}\Bigg[\sum_{P_{X'|Y}:P_{X'}=Q_X} N_{m,\boldsymbol{y}}(P_{X'Y})e^{ng(P_{X'Y})} \\
&\qquad \le \exp\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\}\Bigg] \qquad (540)\\
&\le \mathbb{P}\Bigg[\max_{P_{X'|Y}:P_{X'}=Q_X} N_{m,\boldsymbol{y}}(P_{X'Y})e^{ng(P_{X'Y})} \\
&\qquad \le \exp\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\}\Bigg] \qquad (541)\\
&= \mathbb{P}\Bigg[\bigcap_{P_{X'|Y}:P_{X'}=Q_X} \{N_{m,\boldsymbol{y}}(P)e^{ng(P_{X'Y})} \\
&\qquad \le \exp\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\}\}\Bigg] \qquad (542)\\
&= \mathbb{P}\Bigg[\bigcap_{P_{X'|Y}:P_{X'}=Q_X} \{N_{m,\boldsymbol{y}}(P_{X'Y}) \\
&\qquad \le \exp\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P_{X'Y})\}\Bigg\}\Bigg]. \qquad (543)
\end{aligned}$$

As mentioned above, $N_{m,\boldsymbol{y}}(P_{X'Y})$ is a sum of $M-1$ binary-valued random variables, each has the expectation $e^{-nI(X';Y)}$. However, different from i.i.d. random codebook ensembles, these random variables are correlated.

As [20, Appendix B], we argue that by the definition of $\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})$, there must exist some $P^*_{X'|Y}$ such that for $P^*_{X'Y} \triangleq \hat{P}_{\boldsymbol{y}} \times P^*_{X'|Y}$, $I_{P^*}(X';Y) \le R-\varepsilon$ and $R-\varepsilon-I_{P^*}(X';Y) \ge \alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P^*_{X'Y})$. To see why this is true, assume conversely, that for every $P_{X'|Y}$, which define $P_{X'Y} \triangleq \hat{P}_{\boldsymbol{y}} \times P_{X'|Y}$, either $I_P(X';Y) > R-\varepsilon$ or $I_P(X';Y) > R-\varepsilon$ or $R-I_P(X';Y)-\varepsilon < \alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P_{X'Y})$, which means that for every $P_{X'Y}$,

$$\begin{aligned}
R-\varepsilon &< \max\{I_P(X';Y), I_P(X';Y) \\
&\qquad + \alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P_{X'Y})\} \qquad (544)\\
&= I_P(X';Y) + \big[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P_{X'Y})\big]_+, \qquad (545)
\end{aligned}$$

which implies that for every $P_{X'|Y}$, there exists $t \in [0,1]$ such that

$$\begin{aligned}
R-\varepsilon &< \max\{I_P(X';Y), I_P(X';Y) \\
&\qquad + \alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P_{X'Y})\} \qquad (546)\\
&= I_P(X';Y) + t\big[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P_{X'Y})\big], \qquad (547)
\end{aligned}$$

or equivalently,

$$\begin{aligned}
&\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}}) \\
&> \max_{P_{X'|Y}:P_{X'}=Q_X}\min_{0\le t\le 1} g(P_{X'Y}) + \frac{R-I_P(X';Y)-\varepsilon}{t} \qquad (548)\\
&= \max_{\substack{P_{X'|XY}:P_{X'}=Q_X,\\ I_P(X';Y)\le R-\varepsilon}} \big[g(P_{X'Y})-I_P(X';Y)\big]+R-\varepsilon \qquad (549)\\
&= \alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}}), \qquad (550)
\end{aligned}$$

which is a contradiction.

Now, from (543) and the existence of $P^*_{X'Y}$ as above, it holds that

$$\begin{aligned}
&\mathbb{P}\big[Z_m(\boldsymbol{y}) \le \exp\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\}\big] \\
&\le \mathbb{P}\big[N_{m,\boldsymbol{y}}(P^*_{X'Y}) \le \exp\{n[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P^*_{X'Y})]\}\big]. \qquad (551)
\end{aligned}$$

Different from [11], $N_{\boldsymbol{y}}(P^*_{X'Y})$ is now not the sum of i.i.d. Bernoulli random variables but these random variables are still identically distributed and weakly dependent.

Now, let

$$Z_{m'} \triangleq \mathbb{1}\{(\boldsymbol{X}_{m'},\boldsymbol{y}) \in \mathcal{T}(P^*_{X'Y})\}, \forall m' \in M_- \triangleq [M]\setminus\{m\}, \tag{552}$$

and

$$p \triangleq \mathbb{P}\big[(\boldsymbol{X}_2,\boldsymbol{y}) \in \mathcal{T}(P^*_{X'Y})\big]. \tag{553}$$

Now, let $\nu \in (0,p)$ be chosen such that

$$(M-1)(p-\nu) = \exp\{n[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P^*_{X'Y})]\}. \tag{554}$$

The existence of ν is guaranteed since (554) is equivalent to

$$\begin{aligned}
\nu &= p - \frac{\exp\{n[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P^*_{X'Y})]\}}{M-1} \qquad (555)\\
&\ge p - \frac{\exp\{n[R-\varepsilon-I_{P^*}(X';Y)]\}}{M-1} \qquad (556)\\
&= p - \frac{\exp\{n[R-\varepsilon-I_{P^*}(X';Y)]\}}{\exp(nR)-1} \qquad (557)\\
&\doteq \exp\{-nI_{P^*}(X';Y)\} \\
&\qquad - \exp\{-n(I_{P^*}(X';Y)+\varepsilon)\} > 0, \qquad (558)
\end{aligned}$$

so $\nu \in (0,p)$.

By applying Lemma 19 with $n = M-1$, $X_i = Z_i$, $p = \mathbb{P}\big[(\boldsymbol{X}_2,\boldsymbol{y}) \in \mathcal{T}(P^*_{X'Y})\big]$, and ν satisfying (554), we have

$$\begin{aligned}
&\mathbb{P}\Big[N_{\boldsymbol{y}}(P^*_{X'Y}) \le \exp\{n[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P^*_{X'Y})]\}\Big] \\
&\doteq \mathbb{P}\Big[N_{\boldsymbol{y}}(P^*_{X'Y}) \le \exp\{n[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P^*_{X'Y})]\}\Big] \qquad (559)\\
&\le 2\exp\big(-(M-1)D(p-\nu\|p)\big) \qquad (560)\\
&\doteq \exp(-e^{nR}D(p-\nu\|p)). \qquad (561)
\end{aligned}$$

Now, since $p \doteq \exp(-nI_{P^*}(X';Y))$, from (558), we also have

$$(M-1)\big[(\gamma-1)(p-\nu)\big]$$

$$\dot{\leq} \exp(nR)\left[\left(\frac{1}{1-e^{-n\delta}}-1\right)\exp\left\{-n(I_{P^*}(X';Y)+\varepsilon)\right\}\right] \tag{562}$$

$$\dot{\leq} \frac{e^{-n(\delta+\varepsilon)}}{1-e^{-n\delta}}\exp\left[n(R-I_{P^*}(X';Y))\right]. \tag{563}$$

On the other hand, we have

$$\exp(-e^{nR}D(p-\nu\|p)) = \exp\left\{-e^{nR}D(e^{-an}\|e^{-bn})\right\} \tag{564}$$

where $a \triangleq R+g(P^*_{X'Y})-\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})$ and $b \triangleq I_{P^*}(X';Y)$. It is easy to see that

$$a-b = R+g(P^*_{X'Y})-\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-I_{P^*}(X';Y) \tag{565}$$

$$\geq \varepsilon. \tag{566}$$

Hence, by using the following fact [29, Sec. 6.3]:

$$D(a\|b) \geq a\log\frac{a}{b}+b-a, \tag{567}$$

we have

$$D(e^{-an}\|e^{-bn}) \geq e^{-bn}\left[1+e^{(b-a)n}((b-a)n-1)\right]. \tag{568}$$

Hence, we obtain

$$\exp(-e^{nR}D(p-\nu\|p)) \leq \exp\bigg\{-e^{n(R-I_{P^*}(X';Y))} \times [1-e^{-n\varepsilon}(1+n\varepsilon)]\bigg\}. \tag{569}$$

From (561), (563), and (569), we obtain

$$\mathbb{P}\left[N_{m,\boldsymbol{y}}(P^*_{X'Y}) \leq \exp\left\{n[\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})-g(P^*_{X'Y})]\right\}\right]$$

$$\dot{\leq} \exp\left\{\frac{e^{-n(\delta+\varepsilon)}}{1-e^{-n\delta}}\exp\left[n(R-I_{P^*}(X';Y))\right]\right\} \times\exp\left\{-e^{n(R-I_{P^*}(X';Y))}[1-e^{-n\varepsilon}(1+n\varepsilon)]\right\} \tag{570}$$

$$= \exp\bigg\{-e^{n(R-I_{P^*}(X';Y))} \times\left[1-\frac{e^{-n(\delta+\varepsilon)}}{1-e^{-n\delta}}-e^{-n\varepsilon}(1+n\varepsilon)\right]\bigg\} \tag{571}$$

$$\leq \exp\left\{-e^{n\varepsilon}\left[1-\frac{e^{-n(\delta+\varepsilon)}}{1-e^{-n\delta}}-e^{-n\varepsilon}(1+n\varepsilon)\right]\right\}, \tag{572}$$

where (572) follows from the fact that $I_{P^*}(X';Y) \leq R-\varepsilon$.

From (551) and (572), we obtain

$$\Pr\left[Z_m(\boldsymbol{y}) \leq \exp\left\{n\alpha(R-\varepsilon,\hat{P}_{\boldsymbol{y}})\right\}\right] \dot{\leq} \exp\left\{-e^{n\varepsilon}\left[1-\frac{e^{-n(\delta+\varepsilon)}}{1-e^{-n\delta}}-e^{-n\varepsilon}(1+n\varepsilon)\right]\right\}. \tag{573}$$

This concludes our proof of Lemma 14.

Appendix L
Proof of Lemma 16

The proof is based on [15, Proof of Prep. 5]. However, there are some changes to account for the dependency among the codewords. One such an important change is to replace the Hoeffding's inequality in [15, Proof of Prep. 5] by a generalized version of this inequality in [33].

By using the union bound, we have

$$\mathbb{P}\{\hat{\mathcal{B}}_n(\sigma)\} = \mathbb{P}\left\{\bigcup_{m=1}^{M}\bigcup_{m'\neq m}\bigcup_{\boldsymbol{y}}\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y})\right\} \tag{574}$$

$$\leq \sum_{m=1}^{M}\sum_{m'\neq m}\sum_{\boldsymbol{y}}\mathbb{P}\left\{\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y})\right\}. \tag{575}$$

In addition, for any joint type $P_{XY}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$, let

$$N(P_{XY}) \triangleq \sum_{\tilde{m}\in[M]\setminus\{m,m'\}}\mathbb{1}\{(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{XY})\}, \tag{576}$$

then we also have

$$\mathbb{P}\{\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y})\}$$

$$\doteq \sum_{\substack{P_{XY}:P_X=Q_X,\\ I_P(X;Y)\leq R}}\mathbb{P}\left\{N(P_{XY})\geq e^{n(\beta(R,P_Y)+\sigma-g(P_{XY}))}\right\} + \sum_{\substack{P_{XY}:P_X=Q_X,\\ I_P(X;Y)>R}}\mathbb{P}\left\{N(P_{XY})\geq e^{n(\beta(R,P_Y)+\sigma-g(P_{XY}))}\right\} \tag{577}$$

where (577) follows from [15, Eq. (H.6)].

Now, observe that

$$\mathbb{P}\left\{N(P_{XY})\geq e^{n(\beta(R,P_Y)+\sigma-g(P_{XY}))}\right\}$$

$$\leq \mathbb{P}\left\{N(P_{XY})\geq e^{n(R+\sigma-I_P(X;Y))}\right\} \tag{578}$$

$$= \mathbb{P}\bigg\{\sum_{\tilde{m}\in[M]\setminus\{m,m'\}}\mathbb{1}\{(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{XY})\} \geq e^{n(R+\sigma-I_P(X;Y))}\bigg\} \tag{579}$$

where (578) follows from [15, Eq. (H.9)].

Define a new probability measure Π on $\underbrace{\mathcal{X}^n\times\mathcal{X}^n\cdots\times\mathcal{X}^n}_{M \quad \text{times}}$:

$$\mathbb{P}_\Pi(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M) = \prod_{m=1}^{M}\mathbb{P}(\boldsymbol{X}_m=\boldsymbol{x}_m), \tag{580}$$

for all $(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_M)$.

Note that for any $A\subset[M]\setminus\{m,m'\}$, under the condition (28) we have

$$\mathbb{E}\left[\prod_{\tilde{m}\in A}\mathbb{1}\{(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{XY})\}\right]$$

$$\leq \frac{1}{(1-e^{-n\delta})^{|A|}}\mathbb{E}_\Pi\left[\prod_{\tilde{m}\in A}\mathbb{1}\{(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})\in\mathcal{T}(P_{XY})\}\right] \tag{581}$$

$$= \frac{1}{(1-e^{-n\delta})^{|A|}} \prod_{\tilde{m}\in A} \mathbb{P}\big\{(\tilde{\boldsymbol{X}}_m, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\big\} \tag{582}$$

where (581) follows from the change of measure and Lemma 4.

Now, we have

$$\mathbb{P}\big\{(\tilde{\boldsymbol{X}}_m, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\big\} = \sum_{\tilde{\boldsymbol{x}}_m \in \mathcal{T}(P_{XY}|\boldsymbol{y})} \mathbb{P}(\tilde{\boldsymbol{x}}_m) \tag{583}$$

$$= \sum_{\tilde{\boldsymbol{x}}_m \in \mathcal{T}(P_{XY}|\boldsymbol{y})} \frac{1}{|\mathcal{T}(Q_X)|} \tag{584}$$

$$\doteq e^{-nI_P(X;Y)} \tag{585}$$

where (584) follows from Lemma 3, and (585) follows from [28].

From (582) and (585), we obtain

$$\mathbb{E}\bigg[\prod_{\tilde{m}\in A} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\}\bigg] \dot{\leq} \gamma^{|A|} \tag{586}$$

where

$$\gamma = (1-e^{-n\delta})^{-1} e^{-nI_P(X;Y)}. \tag{587}$$

Hence, if $R \geq I_P(X;Y)$, we have

$$\mathbb{P}\bigg\{\sum_{\tilde{m}\in[M]\setminus\{m,m'\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \geq e^{n(R+\sigma-I_P(X;Y))}\bigg\}$$

$$\dot{\leq} \exp\bigg\{-e^{nR} D\bigg((1-e^{-n\delta})^{-1} e^{\sigma - I_P(X;Y)}\bigg\| (1-e^{-n\delta})^{-1} e^{-nI_P(X;Y)}\bigg)\bigg\} \tag{588}$$

$$\leq \exp\bigg\{-e^{nR}(1-e^{-n\delta})^{-1} e^{-n(I_P(X;Y)-\sigma)} \times \bigg(\log \frac{e^{-n(I_P(X;Y)-\sigma)}}{e^{-nI_P(X;Y)}} - 1\bigg)\bigg\} \tag{589}$$

$$= \exp\bigg\{-(1-e^{-n\delta})^{-1} e^{n(R-I_P(X;Y)+\sigma)}(n\sigma - 1)\bigg\} \tag{590}$$

$$\stackrel{\circ}{\leq} \exp\{-e^{n\sigma}\}, \tag{591}$$

where (588) follows from Lemma 20, (589) follows from the fact that $D(a\|b) \geq a\big(\log\frac{a}{b} - 1\big)$ [29, p. 167], and (591) follows from $R \geq I_P(X;Y)$.

From (579) and (591), we obtain

$$\mathbb{P}\bigg\{N(P_{XY}) \geq e^{n(\beta(R,P_Y)+\sigma-g(P_{XY}))}\bigg\} \stackrel{\circ}{\leq} \exp\{-e^{n\sigma}\} \tag{592}$$

if $I_P(X;Y) \geq R$.

Similarly, for the case $R < I_P(X;Y)$, we have

$$\mathbb{P}\bigg\{N(P_{XY}) \geq e^{n(\beta(R,P_Y)+\sigma-g(P_{XY}))}\bigg\} \leq \mathbb{P}\big\{N(P_{XY}) \geq e^{n\sigma}\big\} \tag{593}$$

$$= \mathbb{P}\bigg\{\sum_{\tilde{m}\in[M]\setminus\{m,m'\}} \mathbb{1}\{(\boldsymbol{X}_{\tilde{m}}, \boldsymbol{y}) \in \mathcal{T}(P_{XY})\} \geq e^{n\sigma}\bigg\}$$

$$\leq \exp\bigg\{-e^{nR} D\big((1-e^{-n\delta})^{-1} e^{-n(R-\sigma)}\big\| (1-e^{-n\delta})^{-1} e^{-nI_P(X;Y)}\big)\bigg\} \tag{594}$$

$$= \exp\bigg\{-(1-e^{-n\delta})^{-1} e^{n\sigma}\big[n(I_P(X;Y) - R + \sigma) - 1\big]\bigg\} \tag{595}$$

$$\stackrel{\circ}{\leq} \exp\{-e^{n\sigma}\}, \tag{596}$$

where (594) is obtained by applying Lemma 20 and the change of measures as the arguments to achieve (591), and (596) follows from the same arguments to achieve (589), and (596) follows from $I_P(X;Y) > R$.

From (577), (592), and (596), we obtain

$$\mathbb{P}\bigg\{\hat{\mathcal{B}}_n(\sigma, m, m', \boldsymbol{y})\bigg\} \stackrel{\circ}{\leq} \exp\{-e^{n\sigma}\}. \tag{597}$$

From (575) and (597), we finally obtain

$$\mathbb{P}\{\hat{\mathcal{B}}_n(\sigma)\} \stackrel{\circ}{\leq} \sum_{m=1}^{M} \sum_{m'\neq m} \sum_{\boldsymbol{y}} \exp\{-e^{n\sigma}\} \tag{598}$$

$$\stackrel{\circ}{=} \exp\{-e^{n\sigma}\}. \tag{599}$$

This concludes our proof of Lemma 16.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

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

[3] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 3–18, Jan. 1965.

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

[5] B. Nakiboğlu, "The sphere packing bound for memoryless channels," *Problems Inf. Transmiss.*, vol. 56, no. 3, pp. 201–244, Jul. 2020.

[6] B. Nakiboğlu, "The Augustin capacity and center," *Problems Inf. Transmiss.*, vol. 55, no. 4, pp. 299–342, Oct. 2019.

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

[8] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 5–12, Jan. 1981.

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

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

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

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

[13] N. Merhav, "Error exponents of typical random trellis codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2067–2077, Apr. 2020.

[14] N. Merhav, "A Lagrange-dual lower bound to the error exponent of the typical random code," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3456–3464, Jun. 2020.

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

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

[17] 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, Apr. 2022.

[18] G. Cocco, A. G. I. Fàbregas, and J. Font-Segura, "Typical error exponents: A dual domain derivation," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 776–793, Feb. 2023.

[19] L. V. Truong, G. Cocco, J. Font-Segura, and A. G. I. Fàbregas, "Concentration properties of random codes," 2022, *arXiv:2203.07853*.

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

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

[22] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1287–1291.

[23] J. Scarlett, A. Martinez, and A. G. I. Fàbregas, "The likelihood decoder: Error exponents and mismatch," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 86–90.

[24] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[25] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of discrete memoryless channels," in *Proc. ISIT*, vol. 77, 1977, p. 107.

[26] E. N. Gilbert, "A comparison of signalling alphabets," *Bell Syst. Tech. J.*, vol. 31, no. 3, pp. 504–522, May 1952.

[27] R. R. Varshamov, "Estimate of the number of signals in error correcting codes," *Doklady Akademii Nauk SSSR*, vol. 117, no. 5, pp. 739–741, 1957.

[28] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[29] N. Merhav, "Statistical physics and information theory," *Found. Trends Commun. Inf. Theory*, vol. 6, nos. 1–2, pp. 1–212, Dec. 2010.

[30] H. Royden and P. Fitzpatrick, *Real Analysis*, 4th ed. London, U.K.: Pearson, 2010.

[31] C. Pelekis and J. Ramon, "Hoeffding's inequality for sums of dependent random variables," *Medit. J. Math.*, vol. 14, p. 243, Nov. 2017.

[32] S. Janson, "New versions of Suen's correlation inequality," *Random Struct. Algorithms*, vol. 13, nos. 3–4, pp. 467–483, Oct. 1998.

[33] R. Impagliazzo and V. Kabanets, "Constructive proofs of concentration bounds," Electron. Colloq. Comput. Complex., Weizmann Inst. Sci., Israel, Tech. Rep. TR10, 2010.

Lan V. Truong (Member, IEEE) received the B.S.E. degree in electronics and telecommunications from the Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam, in 2003, 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 an University Lecturer with the Department of Information Technology Specialization, FPT University, Hanoi. 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, U.K. He is currently a Lecturer (an Assistant Professor) with the School of Mathematics, Statistics and Actuarial Science, University of Essex, U.K. His research interests include information theory, machine learning, data science, and probability.

Albert Guillén i Fàbregas (Fellow, IEEE) received the degree in telecommunications engineering and in electronics engineering from Universitat Politècnica de Catalunya and Politecnico di Torino in 1999 and the Ph.D. degree in communication systems from École Polytechnique Fédérale de Lausanne (EPFL) in 2004.

He is currently a Professor with the Department of Engineering, University of Cambridge, and a part-time Researcher with Universitat Pompeu Fabra. He has held appointments with the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, Universitat Pompeu Fabra, and University of Cambridge, and visiting appointments with EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra (an ICREA Research Professor), University of South Australia, Centrum Wiskunde & Informatica, and Texas A&M University at Qatar. His specific research interests are in the areas of information theory, communication theory, coding theory, and 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 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 WIRELESS COMMUNICATIONS (2007–2011) and IEEE TRANSACTIONS ON INFORMATION THEORY (2013–2020).