Large Deviations Behavior of the Logarithmic Error Probability of Random Codes

Ran Tamir, *Student Member, IEEE*, Neri Merhav, *Fellow, IEEE*, Nir Weinberger, *Member, IEEE*, and Albert Guillén i Fàbregas, *Senior Member, IEEE*

***Abstract*— This work studies the deviations of the error exponent of the constant composition code ensemble around its expectation, known as the error exponent of the typical random code (TRC). In particular, it is shown that the probability of randomly drawing a codebook whose error exponent is smaller than the TRC exponent is exponentially small; upper and lower bounds for this exponent are given, which coincide in some cases. In addition, the probability of randomly drawing a codebook whose error exponent is larger than the TRC exponent is shown to be double–exponentially small; upper and lower bounds to the double–exponential exponent are given. The results suggest that codebooks whose error exponent is larger than the error exponent of the TRC are extremely rare. The key ingredient in the proofs is a new large deviations result of type class enumerators with dependent variables.**

***Index Terms*— Error exponent, expurgated exponent, large deviations, typical random code.**

I. Introduction

RANDOM coding is the most common method to show that the probability of error vanishes for rates below the channel capacity. In 1955, Feinstein [1] proved that, for a sequence of codes of fixed rate and increasing length, the probability of error decays to zero exponentially with the length of the codes, provided that the rate of the code is below the mutual information of the channel. In the same year, Elias [2] derived the random coding and sphere–packing bounds and observed that they exponentially coincide at high rates, for the cases of the binary symmetric channel (BSC) and the binary erasure channel (BEC). Fano [3] derived the random coding exponent, namely,

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

where the expectation is with respect to (w.r.t.) a given ensemble of codes, for the general discrete memoryless channel (DMC). In 1965, Gallager [4] derived $E_{\mathrm{r}}(R)$ in a much simpler way and improved on $E_{\mathrm{r}}(R)$ at low rates by the idea of expurgation.

In random coding analysis, the code is selected at random and remains fixed, and thus, it seems reasonable to study the performance in terms of error exponent of the very chosen code, rather than considering the exponent of the averaged probability of error, as in $E_{\mathrm{r}}(R)$. Therefore, it is natural to ask what would be the error exponent associated with the typical randomly selected code. The error exponent of the typical random code (TRC) is defined as

$$E_{\mathrm{trc}}(R) = \lim_{n\to\infty} \left\{-\tfrac{1}{n}\mathbb{E}\left[\log P_{\mathrm{e}}(\mathcal{C}_n)\right]\right\}. \tag{2}$$

We find the exponent of the TRC to be the more relevant performance metric as it captures the true exponential behavior of the probability of error, as opposed to the random coding error exponent, which is dominated by the relatively poor codes of the ensemble, rather than the channel noise, at relatively low coding rates.

To the best of our knowledge, not much is known on typical random codes. In [5], Barg and Forney considered typical random codes with independently and identically distributed codewords for the BSC with maximum–likelihood (ML) decoding. They also considered typical linear codes. It was shown that at a certain range of low rates, $E_{\mathrm{trc}}(R)$ lies between $E_{\mathrm{r}}(R)$ and the expurgated exponent, $E_{\mathrm{ex}}(R)$. In [6] Nazari *et al*. provided bounds on the error exponent of the TRC for both DMCs and multiple–access channels. In a recent article [7], an exact single–letter expression has been derived for the error exponent of typical, random, constant composition codes, over DMCs, and a wide class of (stochastic) decoders, collectively referred to as the generalized likelihood decoder (GLD), which includes the ML decoder as a special case. For such decoders, the probability of deciding on a given message is proportional to a general exponential function of the joint empirical distribution of the codeword and the received channel output vector. Recently, Merhav has studied

Manuscript received December 20, 2019; revised May 5, 2020; accepted May 8, 2020. Date of publication May 18, 2020; date of current version October 21, 2020. The work of Ran Tamir and Neri Merhav was supported by the Israel Science Foundation (ISF) under Grant 137/18. The work of Nir Weinberger was supported in part by the MIT–Technion Fellowship and in part by the Viterbi Scholarship, Technion. The work of Albert Guillén i Fàbregas was supported in part by the European Research Council through ERC under Grant 725411 and in part by the Spanish Ministry of Economy and Competitiveness under Grant TEC2016-78434-C3-1-R. This article was presented in part at the 2019 IEEE International Symposium on Information Theory. *(Corresponding author: Ran Tamir.)*

Ran Tamir and Neri Merhav are with the Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: rans@campus.technion.ac.il; merhav@ee.technion.ac.il).

Nir Weinberger is with IDSS and LIDS, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: nirw@mit.edu).

Albert Guillén i Fàbregas is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain, also with the Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

Communicated by F. Alajaji, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.2995136

error exponents of TRCs for the colored Gaussian channel [8], typical random trellis codes [9], and a Lagrange–dual lower bound to the TRC exponent [10].

Note that the TRC exponent can be viewed as the limit of the expectation of the random variable

$$\mathsf{E}(\mathcal{C}_n) = -\tfrac{1}{n}\log P_{\text{e}}(\mathcal{C}_n), \tag{3}$$

where $P_{\text{e}}(\mathcal{C}_n)$ is the error probability of a given code $\mathcal{C}_n$, governed by the randomness of the codebook $\mathcal{C}_n$. Having defined this random variable, it is interesting to study, not only its expectation, but also other, more refined, quantities associated with its probability distribution. One of them is the tail behavior, i.e., the large deviations (LD) rate functions. In particular, it is partially implied[1] from [7], that $\mathsf{E}(\mathcal{C}_n)$ concentrates around its expectation, i.e., the error exponent $E_{\text{trc}}(R)$. In this work we prove that $\mathsf{E}(\mathcal{C}_n)$ indeed concentrates around $E_{\text{trc}}(R)$.

In this paper we are interested in probabilities of large fluctuations around $E_{\text{trc}}(R)$. More specifically, we investigate the probability of randomly choosing a *bad* codebook, i.e., a codebook with a relatively small value of $\mathsf{E}(\mathcal{C}_n)$. On the other hand, the probability of randomly drawing a *good* codebook, i.e., a codebook with a relatively large value of $\mathsf{E}(\mathcal{C}_n)$ is of interest as well, since obtaining tight LD bounds is an alternative method to prove upper or lower bounds on the channel reliability function, a long–standing problem.

To the best of our knowledge, the only known bounds on the probability of drawing codebooks with relatively low error exponents are given in [11, Appendix III]. It is proved in [11] that $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) < E_{\text{r}}(R)\}$ is upper bounded by $\exp\{-\exp\{n(R - E_{\text{r}}(R))\}\}$, as long as $R > E_{\text{r}}(R)$, while the entire range of relatively low rates, namely $R \leq E_{\text{r}}(R)$, was hardly considered in [11], and is one of the main topics in the current work. Furthermore, in this paper, we study the deviations of $\mathsf{E}(\mathcal{C}_n)$ w.r.t. its actual expected value $E_{\text{trc}}(R)$, and not as in [11], in which considered deviations w.r.t. $E_{\text{r}}(R)$.

Accordingly, the main purpose of this paper is to study the probabilistic behavior of the tails of $\mathsf{E}(\mathcal{C}_n)$, i.e., to characterize its large deviations properties. For a given $E_0 < E_{\text{trc}}(R)$, we assess the probability $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \leq E_0\}$ and provide exponentially small lower and upper bounds on it, which proves that bad codebooks are rare. More refined questions concerning the lower tail are as follows. Does the probability $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \leq E_0\}$ tend to zero with a finite exponent in the entire range $[0, E_{\text{trc}}(R))$? If not, what is the range of E_0 for which $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \leq E_0\}$ decays faster than exponentially? Indeed, we prove that a *phase transition* occurs in the behavior of this probability, i.e., at some point below $E_{\text{trc}}(R)$, we observe an abrupt change between an ordinary exponential decay to a super–exponential decay. In addition, we consider the probability $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \geq E_0\}$, for $E_0 > E_{\text{trc}}(R)$, and derive double–exponentially small lower and upper bounds on it. We find the largest value E_0, for which $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \geq E_0\}$ is strictly positive, thereby proving the existence of exceptionally good codebooks.

[1] More specifically, for every $\epsilon > 0$, $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \leq E_{\text{trc}}(R) + \epsilon\}$ converges to one exponentially fast as $n \to \infty$.

The remaining part of the paper is organized as follows. In Section 2, we establish notation conventions. In Section 3, we formalize the model, the decoder, LD quantities, and provide some preliminaries. In Section 4, we summarize and discuss the main results, and provide numerical example for the binary z–channel. Sections 5, 6 and 7 include the proofs of our main theorems.

II. Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, realizations will be denoted by the corresponding lower case letters, and their alphabets in calligraphic font. 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 $Q_{XY} = \{Q_{XY}(x,y), x \in \mathcal{X}, y \in \mathcal{Y}\}$, which will often be abbreviated by Q, information measures will be denoted in the conventional manner, but with a subscript Q, that is, $I_Q(X;Y)$ is the mutual information between X and Y, and similarly for other quantities. Logarithms are taken to the natural base. The probability of an event $\mathcal{E}$ will be denoted by $\mathbb{P}\{\mathcal{E}\}$, and the expectation operator will be denoted by $\mathbb{E}[\cdot]$. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathcal{I}\{\mathcal{E}\}$. The notation $[t]_+$ will stand for $\max\{0,t\}$.

For two positive sequences, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ will stand for equality in the exponential scale, that is, $\lim_{n\to\infty}(1/n)\log(a_n/b_n) = 0$. Similarly, $a_n \stackrel{\cdot}{\leq} b_n$ means that $\limsup_{n\to\infty}(1/n)\log(a_n/b_n) \leq 0$, and so on. Accordingly, the notation $a_n \doteq e^{-n\infty}$ means that a_n decays at a super–exponential rate (e.g. double–exponentially).

By the same token, 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 equality in the double–exponential scale, that is,

$$\lim_{n\to\infty} \frac{1}{n}\log\left(\frac{\log b_n}{\log a_n}\right) = 0. \tag{4}$$

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

$$\limsup_{n\to\infty} \frac{1}{n}\log\left(\frac{\log b_n}{\log a_n}\right) \leq 0, \tag{5}$$

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

$$\liminf_{n\to\infty} \frac{1}{n}\log\left(\frac{\log b_n}{\log a_n}\right) \geq 0. \tag{6}$$

The empirical distribution 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 joint empirical distribution of a pair of sequences, denoted by $\hat{P}_{\boldsymbol{xy}}$, is similarly defined. The type class of Q_X, denoted $\mathcal{T}(Q_X)$, is the set of all vectors $\boldsymbol{x} \in \mathcal{X}^n$ with $\hat{P}_{\boldsymbol{x}} = Q_X$. In the same spirit, the joint type class of Q_{XY}, denoted $\mathcal{T}(Q_{XY})$, is the set of all pairs of sequences $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ with $\hat{P}_{\boldsymbol{xy}} = Q_{XY}$.

Throughout the paper, we will make a frequent use of the fact that

$$\sum_{i=1}^{k_n} a_n(i) \doteq \max_{1\leq i\leq k_n} a_n(i) \tag{7}$$

as long as $\{a_n(i)\}$ are nonnegative exponential functions of an integer n and $k_n \doteq 1$. This exponential equivalence will be termed henceforth the *summation–maximization equivalence* (SME). The sequence k_n will represent the number of type classes possible for a given block length n, which is polynomial in n.

III. Problem Formulation

Consider a DMC $W = \{W(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$, where $\mathcal{X}$ and $\mathcal{Y}$ are the finite input and output alphabets, respectively. When the channel is fed with a sequence $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, it produces $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ according to

$$W(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^{n} W(y_t|x_t). \tag{8}$$

Let $\mathcal{C}_n$ be a codebook, i.e., a collection $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$ of $M = e^{nR}$ codewords, n being the block–length and R the coding rate in nats per channel use. When the transmitter wishes to convey a message $m \in \{0, 1, \ldots, M-1\}$, it feeds the channel with $\boldsymbol{x}_m$. We assume that messages are chosen with equal probability. We consider the ensemble of constant composition codes: for a given distribution Q_X over $\mathcal{X}$, all vectors in $\mathcal{C}_n$ are uniformly and independently drawn from the type class $\mathcal{T}(Q_X)$. As in [7], [12], we consider here the GLD, which is a stochastic decoder, that chooses the estimated message $\hat{m}$ according to the following posterior probability mass function, induced by the channel output $\boldsymbol{y}$:

$$\mathbb{P}\left\{\hat{M} = m \middle| \boldsymbol{y}\right\} = \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}}{\sum_{m'=0}^{M-1} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}, \tag{9}$$

where $\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}$ is the empirical distribution of $(\boldsymbol{x}_m, \boldsymbol{y})$, and $g(\cdot)$ is a given continuous, real–valued functional of this empirical distribution. The GLD provides a unified framework which covers several important special cases, e.g., matched likelihood decoding, mismatched decoding, ML decoding, and universal decoding (similarly to the α–decoders described in [13]). In particular, we recover the ML decoder by choosing the decoding metric

$$g(Q_{XY}) = \beta \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} Q_{XY}(x,y) \log W(y|x), \tag{10}$$

and letting $\beta \to \infty$. A more detailed discussion is given in [12].

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

$$P_{\mathrm{e}}(\mathcal{C}_n) = \frac{1}{M} \sum_{m=0}^{M-1} \sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \frac{\sum_{m'\neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\sum_{\tilde{m}=0}^{M-1} \exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}. \tag{11}$$

For the constant composition ensemble, Merhav [7] has derived a single–letter expression for

$$E_{\mathrm{trc}}(R) = \lim_{n\to\infty} \left\{-\tfrac{1}{n}\mathbb{E}\left[\log P_{\mathrm{e}}(\mathcal{C}_n)\right]\right\}. \tag{12}$$

In order to present this expression, we define first a few quantities. Define the set $\mathcal{Q}(Q_X) = \{Q_{XX'} : Q_{X'} = Q_X\}$ and

$$\alpha(R, Q_Y) = \max_{Q_{\tilde{X}|Y}\in\mathcal{S}(Q_X,Q_Y)} \{g(Q_{\tilde{X}Y}) - I_Q(\tilde{X};Y)\} + R, \tag{13}$$

where $\mathcal{S}(Q_X, Q_Y) = \{Q_{\tilde{X}|Y} :\ I_Q(\tilde{X};Y) \leq R,\ Q_{\tilde{X}} = Q_X\}$, as well as

$$\Gamma(Q_{XX'}, R) = \min_{Q_{Y|XX'}} \{D(Q_{Y|X}\|W|Q_X) + I_Q(X';Y|X) + [\max\{g(Q_{XY}), \alpha(R,Q_Y)\} - g(Q_{X'Y})]_+\}, \tag{14}$$

where $D(Q_{Y|X}\|W|Q_X)$ is the conditional divergence between $Q_{Y|X}$ and W, averaged by Q_X:

$$D(Q_{Y|X}||W|Q_X) = \sum_{x\in\mathcal{X}} Q_X(x) \sum_{y\in\mathcal{Y}} Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{W(y|x)}. \tag{15}$$

The TRC error exponent is given by [7][2]

$$E_{\mathrm{trc}}(R) = \min_{\{\mathcal{Q}(Q_X):\ I_Q(X;X')\leq 2R\}} \{\Gamma(Q_{XX'}, R) + I_Q(X;X') - R\}. \tag{16}$$

In the sequel, we prove that the exponent $E_{\mathrm{trc}}(R)$ is the exact value around which the random variable $\mathsf{E}(\mathcal{C}_n)$ concentrates, as was partially implied from the proof in [7, Sec. 5.2]. The expurgated exponent $E_{\mathrm{ex}}(R)$, proved in [12], has exactly the same expression, but with the minimization constraint in (16) $I_Q(X;X') \leq 2R$ replaced by $I_Q(X;X') \leq R$. In case of ML decoding, define

$$a(R, Q_Y) = \max_{Q_{\tilde{X}|Y}\in\mathcal{S}(Q_X,Q_Y)} \mathbb{E}_Q[\log W(Y|\tilde{X})] \tag{17}$$

and the set

$$\mathcal{A}(R) = \{Q_{X'Y|X} :\ I_Q(X;X') \leq 2R,\ Q_{X'} = Q_X,\ \mathbb{E}_Q[\log W(Y|X')] \geq \max\{\mathbb{E}_Q[\log W(Y|X)], a(R,Q_Y)\}\}. \tag{18}$$

Then, (16) particularizes to [7, Sec. 4]

$$E_{\mathrm{trc}}^{\mathrm{ML}}(R) = \min_{Q_{X'Y|X}\in\mathcal{A}(R)} \{D(Q_{Y|X}\|W|Q_X) + I_Q(X,Y;X') - R\}. \tag{19}$$

We are interested in the lower and the upper tails of the distribution of $\mathsf{E}(\mathcal{C}_n)$. The first is

$$\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) \leq E_0\right\}, \quad E_0 < E_{\mathrm{trc}}(R), \tag{20}$$

[2]Note that the expressions of $\alpha(R, Q_Y)$, $\Gamma(Q_{XX'}, R)$, and $E_{\mathrm{trc}}(R)$ are defined in [7] using supremum and infimum. Since all the objective functions involved in the optimization problems defining these terms are continuous and the corresponding feasible sets are compact, these supremum and infimum are in fact achieved by a maxima and minima.

which is the probability of drawing a *bad* codebook. The second one is

$$\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) \geq E_0\right\}, \quad E_0 > E_{\text{trc}}(R), \tag{21}$$

which is the probability of drawing a *good* codebook. Finding exact expressions for (20) and (21) appears to be difficult. We derive lower and upper bounds on both (20) and (21).

IV. Main Results

A. The Lower Tail

In order to present the error exponents of the lower tail, we define the quantities:

$$\beta(R, Q_Y) = \max_{\{Q_{\tilde{X}|Y}:\ Q_{\tilde{X}}=Q_X\}} \{g(Q_{\tilde{X}Y}) + [R - I_Q(\tilde{X};Y)]_+\}, \tag{22}$$

$$\Lambda(Q_{XX'}, R) = \min_{Q_{Y|XX'}} \{D(Q_{Y|X}\|W|Q_X) + I_Q(X';Y|X) + \beta(R, Q_Y) - g(Q_{X'Y})\}, \tag{23}$$

and,

$$\Psi(R, E_0, Q_{XX'}) = \Gamma(Q_{XX'}, R) + R - E_0, \tag{24}$$

$$\Xi(R, E_0, Q_{XX'}) = \Lambda(Q_{XX'}, R) + R - E_0. \tag{25}$$

Also, define the sets

$$\mathcal{L}(R, E_0) = \{Q_{XX'} \in \mathcal{Q}(Q_X) : [2R - I_Q(X;X')]_+ \geq \Psi(R, E_0, Q_{XX'})\}, \tag{26}$$

$$\mathcal{M}(R, E_0) = \{Q_{XX'} \in \mathcal{Q}(Q_X) : [2R - I_Q(X;X')]_+ \geq \Xi(R, E_0, Q_{XX'})\}, \tag{27}$$

and the error exponent functions

$$E_{\text{lt}}^{\text{ub}}(R, E_0) = \min_{Q_{XX'} \in \mathcal{L}(R,E_0)} [I_Q(X;X') - 2R]_+, \tag{28}$$

$$E_{\text{lt}}^{\text{lb}}(R, E_0) = \min_{Q_{XX'} \in \mathcal{M}(R,E_0)} [I_Q(X;X') - 2R]_+. \tag{29}$$

Our first result in this section is the following theorem, which is proved in Section V.

Theorem 1: Consider the ensemble of random constant composition codes $\mathcal{C}_n$ of rate R and composition Q_X. Then,

$$\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) \leq E_0\right\} \stackrel{\cdot}{\leq} \exp\{-n \cdot E_{\text{lt}}^{\text{ub}}(R, E_0)\}. \tag{30}$$

Also,

$$\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) \leq E_0\right\} \stackrel{\cdot}{\geq} \exp\{-n \cdot E_{\text{lt}}^{\text{lb}}(R, E_0)\}. \tag{31}$$

An expression for the special case of ML decoding can be derived, but turns out to be relatively cumbersome, since it consists of a nested optimization problem. Instead, let us recall the result of [14] (see also [15]), which asserts that the probability of error for ordinary likelihood decoding ([12, eq. (3)]) is at most twice the error probability of ML decoding. Hence, it is enough to use the decoding metric $g(Q) = \mathbb{E}_Q[\log W(Y|X)]$ (here and in all of the results later on) in order to study the LD rate functions under the ML decoder. For example, (13) particularizes to

$$\alpha(R, Q_Y) = \max_{Q_{\tilde{X}|Y} \in \mathcal{S}(Q_X, Q_Y)} \{\mathbb{E}_Q[\log W(Y|\tilde{X})] - I_Q(\tilde{X};Y)\} + R, \tag{32}$$

and similarly for $\Gamma(Q_{XX'}, R)$, $\beta(R, Q_Y)$, and $\Lambda(Q_{XX'}, R)$.

We now provide some intuition concerning the term $\Gamma(Q_{XX'}, R)$, which is encountered numerous times in this work. For the true codeword $\boldsymbol{x}_m$ and a competing codeword $\boldsymbol{x}_{m'}$, the term $\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}}, R)$ is, in fact, the exponential rate of decay of the sum

$$\sum_{\boldsymbol{y} \in \mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \exp\left\{-n \cdot [\max\{g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}), \alpha(R, \hat{P}_{\boldsymbol{y}})\} - g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})]_+\right\}, \tag{33}$$

where $g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})$ and $g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})$ are the respective scores of the true and the competing codewords, and where $\alpha(R, \hat{P}_{\boldsymbol{y}})$ represents the highest score among all other incorrect codewords in the codebook[3]. When averaged over all possible channel outputs, this sum yields the overall probability that m' is the decoded message. It follows by the method of types that for a given empirical distribution $\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}}$, there exist some $Q_{Y|XX'}$, such that the most likely channel outputs are those in $\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m, \boldsymbol{x}_{m'})$, and they have the dominant tone in this error event.

In order to characterize the behavior of the error exponent functions (28) and (29), let us first define

$$\tilde{E}(R) = \min_{\{\mathcal{Q}(Q_X):\ I_Q(X;X') \leq 2R\}} \{\Lambda(Q_{XX'}, R) + I_Q(X;X') - R\}. \tag{34}$$

The following proposition is proved in Appendix D.

Proposition 1: $E_{\text{lt}}^{\text{ub}}(R, E_0)$ and $E_{\text{lt}}^{\text{lb}}(R, E_0)$ have the following properties:

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

$$E_0^{\min}(R) = \min_{\mathcal{Q}(Q_X)} \{\Gamma(Q_{XX'}, R) - [2R - I_Q(X;X')]_+\} + R. \tag{35}$$

Note that $\tilde{E}(R)$ is defined similarly as $E_{\text{trc}}(R)$, with $\Lambda(Q_{XX'}, R)$ replacing $\Gamma(Q_{XX'}, R)$. Generally, $\tilde{E}(R) \geq E_{\text{trc}}(R)$, but in some special cases, e.g. the z–channel and the BEC, it can be easily proved that $\tilde{E}(R) = E_{\text{trc}}(R)$, as can be seen in Fig. 3 below. Moreover, since $E_{\text{lt}}^{\text{ub}}(R, E_0)$ is defined similarly as $E_{\text{lt}}^{\text{lb}}(R, E_0)$, also with $\Lambda(Q_{XX'}, R)$ replacing

[3] Also find a more comprehensive discussion on this point in [7, Sec. 4].

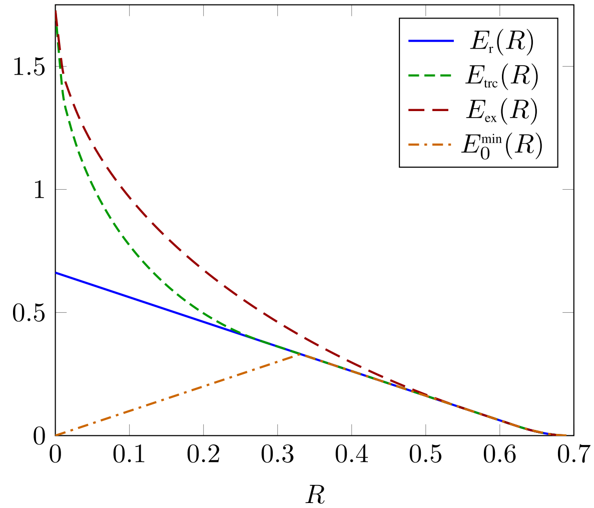


Fig. 1. Various exponents for the z–channel with crossover probability 0.001.

$\Gamma(Q_{XX'}, R)$, it turns out that for the same special cases, $E_{\mathrm{lt}}^{\mathrm{ub}}(R, E_0) = E_{\mathrm{lt}}^{\mathrm{lb}}(R, E_0)$. Hence, we conclude that there exist channels for which $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \leq E_0\}$ has an exponentially tight expression.

Proposition 1 answers the questions we raised in the Introduction. First, it asserts that drawing a codebook for which $\mathsf{E}(\mathcal{C}_n)$ is strictly below the TRC exponent has an exponentially vanishing probability. This implies that only for a small fraction of constant composition codes, $\mathsf{E}(\mathcal{C}_n)$ is significantly lower than the TRC error exponent. Second, the probability that $\mathsf{E}(\mathcal{C}_n)$ falls in the range $(E_0^{\mathrm{min}}(R), E_{\mathrm{trc}}(R))$ tends to zero with a finite exponent, but for $E_0 \in [0, E_0^{\mathrm{min}}(R))$, the probability of $\mathsf{E}(\mathcal{C}_n) \leq E_0$ converges to zero faster than exponentially; these codebooks are extremely rare.

We next describe the behavior of $E_0^{\mathrm{min}}(R)$. Denote by $Q_{XX'}^*(R)$ the minimizer of (35) at rate R, and let R^* be the maximal rate for which $2R \leq I_{Q^*(R)}(X;X')$ holds. On the one hand, for any $R \in [0, R^*]$, the operator $[\cdot]_+$ in (35) is active and $E_0^{\mathrm{min}}(R)$ is given by

$$E_0^{\mathrm{min}}(R) = \min_{\{\mathcal{Q}(Q_X):\ 2R \leq I_Q(X;X')\}} \Gamma(Q_{XX'}, R) + R, \quad (36)$$

which is a monotonically increasing function. On the other hand, if $R \geq R^*$, the operator $[\cdot]_+$ in (35) is neutral and $E_0^{\mathrm{min}}(R)$ coincides with the TRC error exponent $E_{\mathrm{trc}}(R)$. Fig. 1 illustrates the error exponents, as well as $E_0^{\mathrm{min}}(R)$, for the binary z–channel with crossover parameter 0.001, the symmetric input distribution, $Q_X = (1/2, 1/2)$, and the ML decoder. The highest transmission rate is $R \cong 0.685$ [nats/channel use]. As can be seen in Fig. 1, the exponent $E_{\mathrm{trc}}(R)$ lies between $E_{\mathrm{r}}(R)$ and $E_{\mathrm{ex}}(R)$, a fact that was already asserted for a general DMC in [7]. Moreover, $E_{\mathrm{trc}}(R)$ is strictly higher than $E_{\mathrm{r}}(R)$ for relatively low coding rates, and above $R \cong 0.279$ [nats/channel use], they coincide, i.e., the random coding error exponent provides the true exponential behavior of the typical codes in the ensemble. As for $E_0^{\mathrm{min}}(R)$, we observe the following phenomena: First, note that $E_0^{\mathrm{min}}(0) = 0$, which means that all codebooks that have a sub–exponential number of codewords are drawn with a finite exponent. Second, in the range $(0, R^*)$, $E_0^{\mathrm{min}}(R)$ is linear and divides the range $[0, E_{\mathrm{trc}}(R))$ into two intervals; in $(E_0^{\mathrm{min}}(R), E_{\mathrm{trc}}(R))$ – an exponential decay with a finite exponent, and in $[0, E_0^{\mathrm{min}}(R))$ – a super–exponential decay. Third, for rates above R^*, the curves $E_0^{\mathrm{min}}(R)$, $E_{\mathrm{trc}}(R)$, and $E_{\mathrm{r}}(R)$ are all equal. We conclude that for relatively high rates, $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) < E_{\mathrm{trc}}(R)\}$ converges to zero super–exponentially fast, a fact that was already proved in [11, Th. 5].

In order to gain some intuitive insight behind the various types of behavior of $E_{\mathrm{lt}}^{\mathrm{ub}}(R, E_0)$, it is instructive to examine the properties of the type class enumerators,

$$N(Q_{XX'}) \triangleq \sum_{m=0}^{M-1} \sum_{m' \neq m} \mathcal{I}\left\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'})\right\}, \quad (37)$$

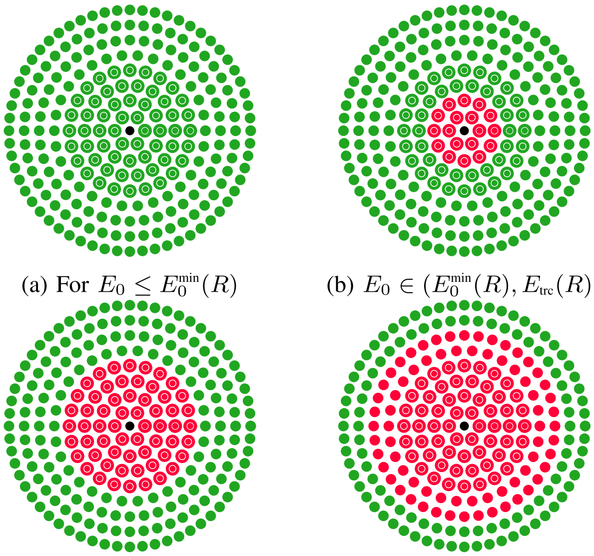
which play a pivotal role in the proofs of the main results of the paper. The summation (37) contains $M(M-1) \doteq e^{n2R}$ terms. Borrowing from the terminology of binomial random variables, we refer to it as the *number of trials* associated with $N(Q_{XX'})$. The expectation of each binary random variable in (37) is given by $\mathbb{P}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'})\} \doteq e^{-nI_Q(X;X')}$, which is referred to as the *success probability*. Unlike its one–dimensional counterpart [16]– [18], $N(Q_{XX'})$ is not a binomial random variable, since its terms are not mutually independent.

We distinguish between two kinds of joint compositions. On the one hand, we have the joint types $Q_{XX'}$ for which $I_Q(X;X') \leq 2R$, i.e., the exponential rate of the number of trials is higher than the negative exponential rate of the success probability. Thus, with overwhelmingly high probability, the respective $N(Q_{XX'})$ will concentrate around its mean, $\exp\{n(2R - I_Q(X;X'))\}$. Such compositions are referred to as *typically populated* (TP) type classes. On the other hand, for $Q_{XX'}$ with $I_Q(X;X') > 2R$, $N(Q_{XX'}) = 0$ with high probability. These compositions are referred to as the *typically empty* (TE) type classes.

For $E_0 \in (E_0^{\mathrm{min}}(R), E_{\mathrm{trc}}(R))$, let us denote the minimizer of $E_{\mathrm{lt}}^{\mathrm{ub}}(R, E_0)$ by $Q_{XX'}^*$. Then, the dominant error event is due to pairs of codewords with joint empirical composition $Q_{XX'}^*$. In this range of exponents, all TP type classes are populated, as well as all TE type classes with $I_Q(X;X') \leq I_{Q^*}(X;X')$. The rest of the TE type classes, those with higher value of $I_Q(X;X')$, are still empty (see Fig. 2b). These are the joint type classes of the "closest" pairs of sequences in $\mathcal{X}^n$, in the sense of high empirical mutual information.

When $E_0 = E_0^{\mathrm{min}}(R)$, the constraint set $\mathcal{L}(R, E_0)$ becomes empty, all TE type classes become populated (see Fig. 2a) and $E_{\mathrm{lt}}^{\mathrm{ub}}(R, E_0)$ jumps to infinity. In some sense, the curve $E_0^{\mathrm{min}}(R)$ exhibits a *phase transition*. When $E_0 > E_0^{\mathrm{min}}(R)$, the minimum "distance" between pairs of codewords is still positive, but when $E_0 \leq E_0^{\mathrm{min}}(R)$, this minimum distance vanishes.

For $E_0 < E_0^{\mathrm{min}}(R)$, the super–exponential behavior of $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \leq E_0\}$ follows from the result of Lemma 5 in Appendix B, which states that $\mathbb{P}\{N(Q_{XX'}) \geq e^{n\epsilon}\}$ tends to zero faster than exponentially for any TE type class. Now, if all TE type classes are populated by exponentially many pairs, then codebooks with exponentially many identical codewords



(a) For $E_0 \le E_0^{\text{min}}(R)$ (b) $E_0 \in (E_0^{\text{min}}(R), E_{\text{trc}}(R))$

(c) Around the $E_{\text{trc}}(R)$ (d) $E_0 \in (E_{\text{trc}}(R), E_{\text{ex}}(R))$

Fig. 2. Typical populations for different E_0 values. The center is the true codeword and each concentric circle around it represents a conditional type class. The radii of the concentric circles represent distances between codewords, which are measured by the empirical conditional entropy (also proportional to the negative empirical mutual information), induced by the joint composition of the codewords. Dots denote the TP type classes and circle–dots represent the TE type classes. TP type classes are the sets of relatively distant codewords; they include all joint compositions $Q_{XX'}$ with $I_Q(X;X') \le 2R$. Red dots/circle–dots mean empty type classes. For larger E_0 values, the minimum distance between codewords increases.

also exist in the range of these low exponents. Consider the set $\mathcal{D}_n = \{\mathcal{C}_n\}$ of codebooks, such that in each one of them, every TE type class is populated by exponentially many pairs of codewords. Obviously, $\mathsf{E}(\mathcal{C}_n) \le E_0^{\text{min}}(R)$ for every $\mathcal{C}_n \in \mathcal{D}_n$, and it turns out that this set has, in fact, a double–exponentially small probability. To see why this is true, consider the following upper bound, which only requires from some $e^{n\epsilon}$ codewords to be identical:

$$\mathbb{P}\{\mathcal{C}_n \in \mathcal{D}_n\} \le \binom{e^{nR}}{e^{n\epsilon}} \cdot \left(\frac{1}{|\mathcal{T}(Q_X)|}\right)^{e^{n\epsilon}} \quad (38)$$

$$\stackrel{\circ}{=} \binom{e^{nR}}{e^{n\epsilon}} \cdot \exp\{-nH_Q(X)e^{n\epsilon}\}. \quad (39)$$

The binomial coefficient is upper–bounded as

$$\binom{e^{nR}}{e^{n\epsilon}} \le \exp\{nRe^{n\epsilon}\}, \quad (40)$$

hence,

$$\mathbb{P}\{\mathcal{C}_n \in \mathcal{D}_n\} \stackrel{\circ}{\le} \exp\{-n(H_Q(X) - R)e^{n\epsilon}\}, \quad (41)$$

which decays double–exponentially fast, since $R < I_Q(X;Y) \le H_Q(X)$.

At last, we prove that a concentration property holds:

Proposition 2: $\mathsf{E}(\mathcal{C}_n)$ concentrates at $E_{\text{trc}}(R)$ as $n \to \infty$.

Proof: On the one hand, it follows by Theorem 1 and Proposition 1 that for every $\epsilon > 0$, $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \le E_{\text{trc}}(R) - \epsilon\} \to 0$, exponentially fast, as $n \to \infty$. On the other hand, the proof in [7, Sec. 5.2] implies that for every $\epsilon > 0$, $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \le E_{\text{trc}}(R) + \epsilon\} \to 1$, also exponentially fast, as $n \to \infty$. Combining these two facts, it follows that $\mathsf{E}(\mathcal{C}_n)$ concentrates at $E_{\text{trc}}(R)$.

B. The Upper Tail

In this subsection, we study the probability $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \ge E_0\}$. On the one hand, we are interested in lower–bounding the probability $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \ge E_0\}$, such that we can assure the existence of good codebooks. On the other hand, we would also like to provide a tight upper bound on this probability, in order to prove that above some critical exponent value, codebooks cease to exist. We begin with a few definitions. Let us define the sets

$$\mathcal{V}(R, E_0) = \{Q_{XX'} \in \mathcal{Q}(Q_X): \; I_Q(X;X') \le 2R, \; \Lambda(Q_{XX'}, R) + I_Q(X;X') - R \le E_0\}, \quad (42)$$

$$\mathcal{U}(R, E_0) = \{Q_{XX'} \in \mathcal{Q}(Q_X): \; I_Q(X;X') \le 2R, \; \Gamma(Q_{XX'}, R) + I_Q(X;X') - R \le E_0\}, \quad (43)$$

and the error exponent functions

$$E_{\text{ut}}^{\text{ub}}(R, E_0) = \max_{Q_{XX'} \in \mathcal{V}(R,E_0)} \min\{2R - I_Q(X;X'), \; E_0 - \Lambda(Q_{XX'}, R) - I_Q(X;X') + R, R\}, \quad (44)$$

$$E_{\text{ut}}^{\text{lb}}(R, E_0) = \max_{Q_{XX'} \in \mathcal{U}(R,E_0)} \{2R - I_Q(X;X')\}. \quad (45)$$

The main result in this subsection is the following theorem.

Theorem 2: Consider the ensemble of random constant composition codes $\mathcal{C}_n$ of rate R and composition Q_X. Then,

$$\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \ge E_0\} \stackrel{\circ}{\le} \exp\{-\exp\{n \cdot E_{\text{ut}}^{\text{ub}}(R, E_0)\}\}. \quad (46)$$

If $E_0 \in (E_{\text{trc}}(R), E_{\text{ex}}(R))$, then

$$\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) \ge E_0\} \stackrel{\circ}{\ge} \exp\{-\exp\{n \cdot E_{\text{ut}}^{\text{lb}}(R, E_0)\}\}. \quad (47)$$

The proofs of (46) and (47) appear in Sections VI and VII, respectively. The double–exponential behavior indicates that the relative number of very good codebooks is extremely small.

The restriction to $(E_{\text{trc}}(R), E_{\text{ex}}(R))$ in the lower bound of Theorem 2 stems from the technical condition of [19, Th. 9], which is equivalent to the one found in the Lovász local lemma [21]. If a large number of events are all independent and each has probability less than 1, then there is a positive probability that none of the events will occur. The Lovász local lemma allows one to slightly relax the independence condition, as long as the events are only "weakly" dependent in some sense. More specifically, referring to the type class enumerator $N(Q_{XX'})$, it turns out that if $I_Q(X;X') > R$, then the binary random variables composing $N(Q_{XX'})$ are only weakly dependent, and the probability $\mathbb{P}\{N(Q_{XX'}) = 0\}$, which appears in the derivation of the lower bound of Theorem 2, can be lower–bounded using the Lovász local lemma by $\exp\{-\exp\{n(2R - I_Q(X;X'))\}\}$. Otherwise, when $I_Q(X;X') < R$, this probability is very small, but it cannot be lower–bounded by the Lovász local lemma, since its condition is not met. In our setting, the condition of the local lemma is met, as long as the number of codewords is not too high, which results in an upper bound on E_0, given by $E_{\text{ex}}(R)$.

In order to characterize the behavior of the error exponent functions (44) and (45), we provide the following proposition, which is proved in Appendix E.

Proposition 3: $E_{\text{ut}}^{\text{ub}}(R,E_0)$ and $E_{\text{ut}}^{\text{lb}}(R,E_0)$ have the following properties:

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

Recall that for the typical code, i.e., any code with $\mathsf{E}(\mathcal{C}_n) \approx E_{\text{trc}}(R)$, all TP type classes are populated and all TE type classes are empty (see Fig. 2c). Now, for any E_0 in the range $(E_{\text{trc}}(R), E_{\text{ex}}(R))$, all TE type classes are still empty, but now, also all TP type classes that are associated with the set $\mathcal{U}(R,E_0)$ are also empty (see Fig. 2d). The dominant error event in these codebooks is caused by relatively distant pairs of codewords that have a joint composition $Q_{XX'}^*$, which is the maximizer of (45). We conclude that $E_{\text{trc}}(R)$ exhibits a phase transition in the E_0 axis. Below the $E_{\text{trc}}(R)$ curve, TE type classes become populated, and above it, TP type classes become empty.

When E_0 reaches $E_{\text{ex}}(R)$, the set $\mathcal{U}(R,E_{\text{ex}})$ is a subset of $\tilde{\mathcal{U}}(R) = \{Q_{XX'} \in \mathcal{Q}(Q_X) : \; R < I_Q(X;X') \le 2R\}$, and thus

$$E_{\text{ut}}^{\text{lb}}(R,E_{\text{ex}}) = \max_{\mathcal{U}(R,E_{\text{ex}})} \{2R - I_Q(X;X')\} \tag{48}$$

$$\le \max_{\tilde{\mathcal{U}}(R)} \{2R - I_Q(X;X')\} = R. \tag{49}$$

It means that the lower bound of Theorem 2 is at least as high as the probability of any codebook in the ensemble, given by $\overset{\circ}{=} \exp\{-nH_Q(X)e^{nR}\}$, which implies the existence of codebooks with $\mathsf{E}(\mathcal{C}_n) \approx E_{\text{ex}}(R)$. We have the following corollary, which is proved in Appendix F.

Corollary 1: If $E_0 < E_{\text{ex}}(R)$, then there exists at least one code with $\mathsf{E}(\mathcal{C}_n) \ge E_0$.

Fig. 3 illustrates the upper tail exponents (44) and (45) for the binary z–channel with crossover parameter 0.001, rate $R = 0.2$, the symmetric input distribution, $Q_X = (1/2, 1/2)$, and the ML decoder. Due to the restriction in the lower bound of Theorem 2, note that $E_{\text{ut}}^{\text{lb}}(R,E_0)$ is applicable as long as $0 \le E_{\text{ut}}^{\text{lb}}(R,E_0) \le R$, while $E_{\text{ut}}^{\text{ub}}(R,E_0)$ is applicable for any E_0, but is truncated to R for relatively high E_0. The lowest E_0 for which $E_{\text{ut}}^{\text{ub}}(R,E_0) = R$ is approximately 0.873, which is strictly lower than the straight–line bound $E_{\text{sl}}(R) \approx 1.122$, but the truncation[4] to R prevents[5] us from deducing a tighter upper bound to the reliability function. In the entire range $(E_{\text{trc}}(R), E_{\text{ex}}(R))$, both $E_{\text{ut}}^{\text{lb}}(R,E_0)$ and $E_{\text{ut}}^{\text{ub}}(R,E_0)$ are strictly positive, such that the lower and the upper bounds on the probability of the upper tail are double–exponentially small.

[4]We conjecture that this truncation to R is artificial, and can be removed by deriving tighter LD bounds. More specifically, a tighter version of Fact 1 (Appendix A), which may lead to a tighter result in Lemma 2 (Appendix B), which, in turn, may provide a tighter upper bound in Theorem 2.

[5]Had the double–exponential rate of the upper bound strictly bigger than R, we were able to conclude the absentee of codebooks with error exponents above some threshold.

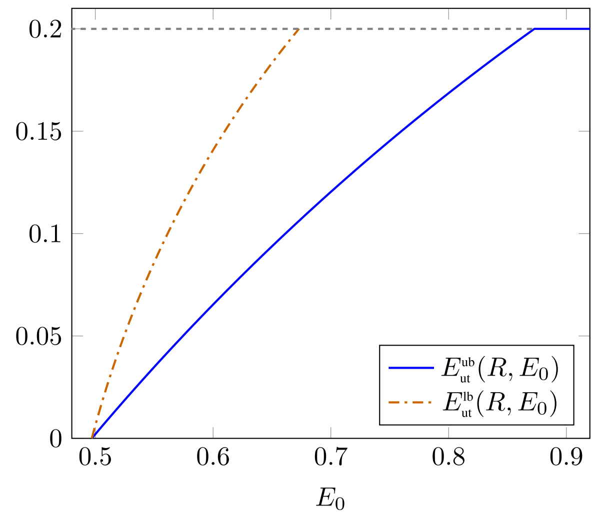


Fig. 3. Upper tail double–exponential rate functions for the z–channel with crossover probability 0.001 and $R = 0.2$.

V. Proof of Theorem 1

A. An Upper Bound on the Probability of the Lower Tail

Let $\mathcal{C}_n$ be a constant composition code of rate R and blocklength n and let $E_0 > 0$ be given. Then,

$$\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \le E_0\right\}$$

$$= \mathbb{P}\left\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\sum_{\tilde{m}}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}} \ge e^{-n\cdot E_0}\right\}. \tag{50}$$

Let

$$Z_m(\boldsymbol{y}) = \sum_{\tilde{m}\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}, \tag{51}$$

fix $\epsilon > 0$ arbitrarily small, and for every $\boldsymbol{y} \in \mathcal{Y}^n$, define the set

$$\mathcal{B}_\epsilon(m,\boldsymbol{y}) = \left\{\mathcal{C}_n : \; Z_m(\boldsymbol{y}) \le \exp\{n\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})\}\right\}. \tag{52}$$

Following the result of [12, Appendix B], we know that, considering the ensemble of randomly selected constant composition codes of type Q_X,

$$\mathbb{P}\{\mathcal{B}_\epsilon(m,\boldsymbol{y})\} \le \exp\{-e^{n\epsilon} + n\epsilon + 1\}, \tag{53}$$

for every $m \in \{0,1,\ldots,M-1\}$ and $\boldsymbol{y} \in \mathcal{Y}^n$, and so, by the union bound,

$$\mathbb{P}\left\{\bigcup_{m=0}^{M-1}\bigcup_{\boldsymbol{y}\in\mathcal{Y}^n}\mathcal{B}_\epsilon(m,\boldsymbol{y})\right\} \triangleq \mathbb{P}\{\mathcal{B}_\epsilon\} \tag{54}$$

$$\le \sum_{m=0}^{M-1}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\{\mathcal{B}_\epsilon(m,\boldsymbol{y})\} \tag{55}$$

$$\leq \sum_{m=0}^{M-1} \sum_{\boldsymbol{y}\in\mathcal{Y}^n} \exp\{-e^{n\epsilon}+n\epsilon+1\} \tag{56}$$
$$= e^{nR}\cdot|\mathcal{Y}|^n\cdot\exp\{-e^{n\epsilon}+n\epsilon+1\}, \tag{57}$$

which still decays double–exponentially fast. Thus,

$$\begin{aligned}
&\mathbb{P}\left\{-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)\leq E_0\right\}\\
&= \mathbb{P}\left\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m)\right.\\
&\qquad\left.\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+Z_m(\boldsymbol{y})}\geq e^{-n\cdot E_0}\right\} &(58)\\
&= \mathbb{P}\left\{\mathcal{C}_n\in\mathcal{B}_\epsilon^{\mathrm{c}}, \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m)\right.\\
&\qquad\left.\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+Z_m(\boldsymbol{y})}\geq e^{-n\cdot E_0}\right\}\\
&\quad+\mathbb{P}\left\{\mathcal{C}_n\in\mathcal{B}_\epsilon, \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m)\right.\\
&\qquad\left.\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+Z_m(\boldsymbol{y})}\geq e^{-n\cdot E_0}\right\} &(59)\\
&\leq \mathbb{P}\left\{\mathcal{C}_n\in\mathcal{B}_\epsilon^{\mathrm{c}}, \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m)\right.\\
&\qquad\times\min\left\{1,\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{n\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})\}}\right\}\\
&\qquad\left.\geq e^{-n\cdot E_0}\right\}+\mathbb{P}\{\mathcal{C}_n\in\mathcal{B}_\epsilon\} &(60)\\
&\doteq \mathbb{P}\left\{\mathcal{C}_n\in\mathcal{B}_\epsilon^{\mathrm{c}}, \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m)\right.\\
&\qquad\times\exp\left\{-n\cdot[\max\{g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}),\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})\}\right.\\
&\qquad\left.\left.-g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})]_+\right\}\geq e^{-n\cdot E_0}\right\}+\mathbb{P}\{\mathcal{C}_n\in\mathcal{B}_\epsilon\} &(61)\\
&\doteq \mathbb{P}\left\{\mathcal{C}_n\in\mathcal{B}_\epsilon^{\mathrm{c}}, \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\exp\{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R-\epsilon)\}\right.\\
&\qquad\left.\geq e^{-n\cdot E_0}\right\} &(62)\\
&\leq \mathbb{P}\left\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m} e^{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R-\epsilon)}\geq e^{-n\cdot E_0}\right\}, &(63)
\end{aligned}$$

where in (60), the inner terms in the first expression of (59) were upper–bounded according to (52) as well as the trivial upper bound of one, and the indicators of the second summand were trivially upper–bounded by one. In (61), we used the SME (7). In (62), the inner–most sum over $\boldsymbol{y}\in\mathcal{Y}^n$ was evaluated using the method of types, with the functional $\Gamma(Q_{XX'},R)$ defined in (14) (see [12, Sec. 5] for more details), and the fact that $\mathbb{P}\{\mathcal{B}_\epsilon\}$ is double–exponentially small was used. One of the difficulties in the statistical analysis of $N(Q_{XX'})$ (37) is that it is the sum of *dependent*[6] (though pairwise independent) binary random variables. This is different from the more commonly encountered type class enumerators (see, e.g., [16], [17], [18]), which are sums of *independent* binary random variables. Hence, existing results concerning the LD for type class enumerators of independent variables are not applicable, and thus, more refined tools from LD theory are required, like those of [19], that will allow us to handle dependency between terms[7]. In spite of the statistical dependencies, it turns out, that the LD behavior of $N(Q_{XX'})$ and the ordinary type class enumerators are the same. This can be seen in the following theorem, which is proved in Appendix B.

Theorem 3: For any $s\in\mathbb{R}$,

$$\mathbb{P}\{N(Q_{XX'})\geq e^{ns}\}\doteq e^{-n\cdot E(R,Q,s)}, \tag{64}$$

where,

$$E(R,Q,s)=\begin{cases}[I_Q(X;X')-2R]_+ & [2R-I_Q(X;X')]_+\geq s\\ \infty & [2R-I_Q(X;X')]_+<s\end{cases}. \tag{65}$$

Then, we rewrite (63) in terms of the enumerators $N(Q_{XX'})$ and get

$$\begin{aligned}
&\mathbb{P}\left\{-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)\leq E_0\right\}\\
&\stackrel{\cdot}{\leq}\mathbb{P}\left\{\sum_{Q_{XX'}\in\mathcal{Q}(Q_X)} N(Q_{XX'})\exp\{-n\cdot\Gamma(Q_{XX'},R-\epsilon)\}\right.\\
&\qquad\left.\geq e^{n\cdot(R-E_0)}\right\} &(66)\\
&\doteq\mathbb{P}\left\{\max_{Q_{XX'}\in\mathcal{Q}(Q_X)} N(Q_{XX'})\exp\{-n\cdot\Gamma(Q_{XX'},R-\epsilon)\}\right.\\
&\qquad\left.\geq e^{n\cdot(R-E_0)}\right\} &(67)\\
&=\mathbb{P}\left\{\bigcup_{Q_{XX'}\in\mathcal{Q}(Q_X)} N(Q_{XX'})\exp\{-n\cdot\Gamma(Q_{XX'},R-\epsilon)\}\right.\\
&\qquad\left.\geq e^{n\cdot(R-E_0)}\right\} &(68)\\
&\doteq\sum_{Q_{XX'}\in\mathcal{Q}(Q_X)}\mathbb{P}\left\{N(Q_{XX'})\exp\{-n\cdot\Gamma(Q_{XX'},R-\epsilon)\}\right.\\
&\qquad\left.\geq e^{n\cdot(R-E_0)}\right\} &(69)\\
&\doteq\max_{Q_{XX'}\in\mathcal{Q}(Q_X)}\mathbb{P}\left\{N(Q_{XX'})\right.\\
&\qquad\left.\geq\exp\{n\cdot(\Psi(R-\epsilon,E_0,Q_{XX'})+\epsilon)\}\right\}, &(70)
\end{aligned}$$

where the steps to (67) and (70) are due to the SME of (7). Define the set $\mathcal{S}_\epsilon(R,E_0)=\{Q_{XX'}:\ [2R-I_Q(X;X')]_+\geq\Psi(R-\epsilon,E_0,Q_{XX'})+\epsilon\}$. Thanks to Theorem 3, the last

[6]This dependence can be demonstrated by the following extreme example. Let Q_X be uniform over $\mathcal{X}$ and let $Q_{XX'}(x,x')=1/|\mathcal{X}|$ whenever $x=x'$ and $Q_{XX'}(x,x')=0$ otherwise. Then, without any prior knowledge, for every $m'\neq m$, $\mathbb{P}\{\boldsymbol{X}_m=\boldsymbol{X}_{m'}\}=\mathbb{P}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(Q_{XX'})\}\doteq\exp\{-nI_Q(X;X')\}$, where $I_Q(X;X')=\log|\mathcal{X}|$. Now, conditioned on $\boldsymbol{X}_0=\boldsymbol{X}_1$ and $\boldsymbol{X}_1=\boldsymbol{X}_2$, it holds that $\boldsymbol{X}_0=\boldsymbol{X}_2$ with probability 1.

[7]Also refer to [20, Sec. IV–C], where bounds from [19] were used to handle weak dependencies in joint types.

expression decays exponentially with rate $E_{\text{lt}}^{\text{ub}}(R, E_0, \epsilon)$, which is given by

$$E_{\text{lt}}^{\text{ub}}(R, E_0, \epsilon) = \min_{Q_{XX'} \in \mathcal{Q}(Q_X)} \begin{cases} [I_Q(X;X') - 2R]_+ & Q_{XX'} \in \mathcal{S}_\epsilon(R, E_0) \\ \infty & Q_{XX'} \notin \mathcal{S}_\epsilon(R, E_0) \end{cases} \quad (71)$$

$$= \min_{Q_{XX'} \in \mathcal{Q}(Q_X) \cap \mathcal{S}_\epsilon(R, E_0)} [I_Q(X;X') - 2R]_+ , \quad (72)$$

with the convention that the minimum over an empty set is defined as infinity. Due to the arbitrariness of $\epsilon > 0$, it follows that

$$\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \le E_0\right\} \dot{\le} \exp\{-n \cdot E_{\text{lt}}^{\text{ub}}(R, E_0)\}, \quad (73)$$

which proves the upper bound of Theorem 1.

B. A Lower Bound on the Probability of the Lower Tail

For a given m, $m' \neq m$, and $\boldsymbol{y} \in \mathcal{Y}^n$, define

$$Z_{mm'}(\boldsymbol{y}) = \sum_{\tilde{m} \in \{0,1,\ldots,M-1\} \setminus \{m,m'\}} \exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}. \quad (74)$$

Let $\sigma > 0$ and define the set

$$\hat{\mathcal{B}}_n(\sigma, m, m', \boldsymbol{y}) = \left\{\mathcal{C}_n : \; Z_{mm'}(\boldsymbol{y}) \ge \exp\{n \cdot (\beta(R, \hat{P}_{\boldsymbol{y}}) + \sigma)\}\right\}, \quad (75)$$

and its complement $\hat{\mathcal{G}}_n(\sigma, m, m', \boldsymbol{y})$, where $\beta(R, Q_Y)$ is defined as in (22). Let

$$\hat{\mathcal{B}}_n(\sigma) = \bigcup_{m=0}^{M-1} \bigcup_{m' \neq m} \bigcup_{\boldsymbol{y} \in \mathcal{Y}^n} \hat{\mathcal{B}}_n(\sigma, m, m', \boldsymbol{y}), \quad (76)$$

and

$$\hat{\mathcal{G}}_n(\sigma) = \hat{\mathcal{B}}_n^{\text{c}}(\sigma). \quad (77)$$

Let $\epsilon > 0$ be arbitrary and define

$$\tilde{\Lambda}(Q_{XX'}, R, \epsilon) = \min_{Q_{Y|XX'}} \{D(Q_{Y|X} \| W | Q_X) + I_Q(X';Y|X) + [\max\{g(Q_{XY}), \beta(R, Q_Y) + \epsilon\} - g(Q_{X'Y})]_+\}. \quad (78)$$

We get the following

$$\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \le E_0\right\}$$

$$= \mathbb{P}\left\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m' \neq m}\sum_{\boldsymbol{y} \in \mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times \frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})} + e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})} + Z_{mm'}(\boldsymbol{y})} \ge e^{-n \cdot E_0}\right\} \quad (79)$$

$$\ge \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\epsilon), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m' \neq m}\sum_{\boldsymbol{y} \in \mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times \frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})} + e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})} + Z_{mm'}(\boldsymbol{y})} \ge e^{-n \cdot E_0}\right\} \quad (80)$$

$$\ge \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\epsilon), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m' \neq m}\sum_{\boldsymbol{y} \in \mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times \frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})} + e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})} + e^{n \cdot [\beta(R, \hat{P}_{\boldsymbol{y}}) + \epsilon]}} \ge e^{-n \cdot E_0}\right\} \quad (81)$$

$$\doteq \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\epsilon), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m' \neq m}\sum_{\boldsymbol{y} \in \mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times e^{n \cdot [\max\{g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}), \beta(R, \hat{P}_{\boldsymbol{y}}) + \epsilon\} - g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})]_+} \ge e^{-n \cdot E_0}\right\} \quad (82)$$

$$\doteq \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\epsilon), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m' \neq m} e^{-n \cdot \tilde{\Lambda}(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}}, R, \epsilon)} \ge e^{-n \cdot E_0}\right\} \quad (83)$$

$$= \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\epsilon), \sum_{Q_{XX'} \in \mathcal{Q}(Q_X)} N(Q_{XX'}) \cdot e^{-n \cdot \tilde{\Lambda}(Q_{XX'}, R, \epsilon)} \ge e^{n \cdot (R - E_0)}\right\}, \quad (84)$$

where (79) follows from the definitions of the probability of error and $Z_{mm'}(\boldsymbol{y})$ in (11) and (74), respectively. In (80), we lower–bounded by intersecting with the event $\mathcal{C}_n \in \hat{\mathcal{G}}_n(\epsilon)$. In (81), the definition of the set $\hat{\mathcal{G}}_n(\cdot)$ in (77) was used, in (82), the exponential equivalence $e^{nB}/(e^{nA}+e^{nB}+e^{nC}) \doteq \exp\{-n \cdot [\max\{A, C\} - B]_+\}$, in (83), the method of types and the definition of $\tilde{\Lambda}(Q_{XX'}, R, \epsilon)$ in (78), and in (84), the definition of the type class enumerators $N(Q_{XX'})$ in (37).

Next, we simplify the expression of $\tilde{\Lambda}(Q_{XX'}, R, \epsilon)$. First, note that for any $\hat{Q}_{XY}$ with marginals Q_X and Q_Y

$$\beta(R, Q_Y) = \max_{\{Q_{\tilde{X}|Y} : \; Q_{\tilde{X}} = Q_X\}} \{g(Q_{\tilde{X}Y}) + [R - I_Q(\tilde{X};Y)]_+\} \quad (85)$$

$$\ge \max_{\{Q_{\tilde{X}|Y} : \; Q_{\tilde{X}} = Q_X\}} g(Q_{\tilde{X}Y}) \quad (86)$$

$$\ge g(\hat{Q}_{XY}). \quad (87)$$

Then,

$$\tilde{\Lambda}(Q_{XX'}, R, \epsilon) = \min_{Q_{Y|XX'}} \{D(Q_{Y|X} \| W | Q_X) + I_Q(X';Y|X) + [\max\{g(Q_{XY}), \beta(R, Q_Y) + \epsilon\} - g(Q_{X'Y})]_+\} \quad (88)$$

$$= \min_{Q_{Y|XX'}} \{D(Q_{Y|X} \| W | Q_X) + I_Q(X';Y|X) + [\beta(R, Q_Y) + \epsilon - g(Q_{X'Y})]_+\} \quad (89)$$

$$= \min_{Q_{Y|XX'}} \{D(Q_{Y|X} \| W | Q_X) + I_Q(X';Y|X) + \beta(R, Q_Y) - g(Q_{X'Y}) + \epsilon\} \quad (90)$$

$$= \Lambda(Q_{XX'}, R) + \epsilon, \quad (91)$$

where (89) is due to $\beta(R, Q_Y) \ge g(Q_{XY})$, (90) is because $\beta(R, Q_Y) \ge g(Q_{X'Y})$, and (91) follows the definition in (23).

Let us now define

$$\mathcal{G}_0 = \Bigg\{\mathcal{C}_n : \sum_{Q_{XX'}\in\mathcal{Q}(Q_X)} N(Q_{XX'}) \cdot \exp\{-n\cdot(\Lambda(Q_{XX'},R)+\epsilon)\} \geq e^{n\cdot(R-E_0)}\Bigg\}, \quad (92)$$

such that, continuing from (84):

$$\begin{aligned}
&\mathbb{P}\left\{-\frac{1}{n}\log P_e(\mathcal{C}_n) \leq E_0\right\} \\
&\dot{\geq} \mathbb{P}\left\{\hat{\mathcal{G}}_n(\epsilon)\cap\mathcal{G}_0\right\} && (93)\\
&= \mathbb{P}\left\{\bigcap_{m=0}^{M-1}\bigcap_{m'\neq m}\bigcap_{\boldsymbol{y}\in\mathcal{Y}^n}\hat{\mathcal{G}}_n(\epsilon,m,m',\boldsymbol{y})\middle|\mathcal{G}_0\right\}\cdot\mathbb{P}\{\mathcal{G}_0\} && (94)\\
&= \left(1-\mathbb{P}\left\{\bigcup_{m=0}^{M-1}\bigcup_{m'\neq m}\bigcup_{\boldsymbol{y}\in\mathcal{Y}^n}\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y})\middle|\mathcal{G}_0\right\}\right)\cdot\mathbb{P}\{\mathcal{G}_0\} && (95)\\
&\geq \left(1-\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\left\{\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y})\middle|\mathcal{G}_0\right\}\right)\cdot\mathbb{P}\{\mathcal{G}_0\} && (96)\\
&= \mathbb{P}\{\mathcal{G}_0\} - \sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\left\{\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y})\cap\mathcal{G}_0\right\}. && (97)
\end{aligned}$$

Assessing $\mathbb{P}\{\mathcal{G}_0\}$ *in* (97)*:* Now,

$$\begin{aligned}
&\mathbb{P}\{\mathcal{G}_0\} \\
&= \mathbb{P}\left\{\sum_{Q_{XX'}\in\mathcal{Q}(Q_X)} N(Q_{XX'})\cdot e^{-n\cdot(\Lambda(Q_{XX'},R)+\epsilon)} \geq e^{n\cdot(R-E_0)}\right\} && (98)\\
&\doteq \sum_{Q_{XX'}\in\mathcal{Q}(Q_X)}\mathbb{P}\left\{N(Q_{XX'})\geq e^{n\cdot(\Lambda(Q_{XX'},R)+R-E_0+\epsilon)}\right\} && (99)\\
&\doteq \max_{Q_{XX'}\in\mathcal{Q}(Q_X)}\mathbb{P}\left\{N(Q_{XX'})\geq e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)}\right\}, && (100)
\end{aligned}$$

where (99) and (100) follow by the SME and are similar to the steps between (66)–(70). Define the set $\mathcal{S}'_\epsilon(R,E_0) = \{Q_{XX'} : [2R - I_Q(X;X')]_+ \geq \Xi(R,E_0,Q_{XX'}) + \epsilon\}$. Thanks to Theorem 3, the last expression decays exponentially with rate $E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon)$, which is given by

$$\begin{aligned}
&E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon) \\
&= \min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\begin{cases}[I_Q(X;X')-2R]_+ & Q_{XX'}\in\mathcal{S}'_\epsilon(R,E_0)\\ \infty & Q_{XX'}\notin\mathcal{S}'_\epsilon(R,E_0)\end{cases} && (101)\\
&= \min_{Q_{XX'}\in\mathcal{Q}(Q_X)\cap\mathcal{S}'_\epsilon(R,E_0)}[I_Q(X;X')-2R]_+, && (102)
\end{aligned}$$

and thus

$$\mathbb{P}\{\mathcal{G}_0\} \doteq \exp\{-n\cdot E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon)\}. \quad (103)$$

Upper–bounding $\mathbb{P}\{\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y})\cap\mathcal{G}_0\}$ *in* (97)*:* Define the type class enumerator

$$N_{\boldsymbol{y}}(Q_{XY}) = \sum_{m=0}^{M-1}\mathcal{I}\{(\boldsymbol{X}_m,\boldsymbol{y})\in\mathcal{T}(Q_{XY})\}. \quad (104)$$

Then, we have the following

$$\begin{aligned}
&\mathbb{P}\{\hat{\mathcal{B}}_n(\epsilon,\hat{m},\ddot{m},\boldsymbol{y})\cap\mathcal{G}_0\} \\
&= \mathbb{P}\Bigg\{\sum_{\tilde{m}\in\{0,1,\ldots,M-1\}\setminus\{\hat{m},\ddot{m}\}} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}}\boldsymbol{y}})} \geq e^{n\cdot(\beta(R,\hat{P}_{\boldsymbol{y}})+\epsilon)},\\
&\qquad \sum_{m=0}^{M-1}\sum_{m'\neq m} e^{-n\cdot(\Lambda(\hat{P}_{\boldsymbol{X}_m\boldsymbol{X}_{m'}},R)+\epsilon)} \geq e^{n\cdot(R-E_0)}\Bigg\} && (105)\\
&\leq \mathbb{P}\Bigg\{\sum_{\tilde{m}\in\{0,1,\ldots,M-1\}} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}}\boldsymbol{y}})} \geq e^{n\cdot(\beta(R,\hat{P}_{\boldsymbol{y}})+\epsilon)},\\
&\qquad \sum_{m=0}^{M-1}\sum_{m'\neq m} e^{-n\cdot(\Lambda(\hat{P}_{\boldsymbol{X}_m\boldsymbol{X}_{m'}},R)+\epsilon)} \geq e^{n\cdot(R-E_0)}\Bigg\} && (106)\\
&= \mathbb{P}\Bigg\{\sum_{Q_{XY}} N_{\boldsymbol{y}}(Q_{XY})e^{ng(Q_{XY})} \geq e^{n\cdot(\beta(R,\hat{P}_{\boldsymbol{y}})+\epsilon)},\\
&\qquad \sum_{Q_{XX'}} N(Q_{XX'})e^{-n\cdot(\Lambda(Q_{XX'},R)+\epsilon)} \geq e^{n\cdot(R-E_0)}\Bigg\} && (107)\\
&\doteq \mathbb{P}\Bigg\{\bigcup_{Q_{XY}}\left\{N_{\boldsymbol{y}}(Q_{XY}) \geq e^{n\cdot(\beta(R,\hat{P}_{\boldsymbol{y}})-g(Q_{XY})+\epsilon)}\right\},\\
&\qquad \bigcup_{Q_{XX'}}\left\{N(Q_{XX'}) \geq e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)}\right\}\Bigg\} && (108)\\
&\doteq \sum_{Q_{XY}}\sum_{Q_{XX'}}\mathbb{P}\Big\{N_{\boldsymbol{y}}(Q_{XY})^l \geq e^{n\cdot(\beta(R,\hat{P}_{\boldsymbol{y}})-g(Q_{XY})+\epsilon)\cdot l},\\
&\qquad N(Q_{XX'})^k \geq e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k}\Big\} && (109)\\
&\doteq \max_{Q_{XY}}\max_{Q_{XX'}}\mathbb{P}\Big\{N_{\boldsymbol{y}}(Q_{XY})^l \geq e^{n\cdot(\beta(R,\hat{P}_{\boldsymbol{y}})-g(Q_{XY})+\epsilon)\cdot l},\\
&\qquad N(Q_{XX'})^k \geq e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k}\Big\} && (110)\\
&\leq \max_{Q_{XY}}\max_{Q_{XX'}}\mathbb{P}\Big\{N_{\boldsymbol{y}}(Q_{XY})^l\cdot N(Q_{XX'})^k\\
&\qquad \geq e^{n\cdot(\beta(R,\hat{P}_{\boldsymbol{y}})-g(Q_{XY})+\epsilon)\cdot l}\cdot e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k}\Big\} && (111)\\
&\leq \max_{Q_{XY}}\max_{Q_{XX'}}\mathbb{P}\Big\{N_{\boldsymbol{y}}(Q_{XY})^l\cdot N(Q_{XX'})^k\\
&\qquad \geq e^{n\cdot([R-I_Q(X;Y)]_++\epsilon)\cdot l}\cdot e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k}\Big\} && (112)\\
&\leq \max_{Q_{XY}}\max_{Q_{XX'}}\frac{\mathbb{E}\left[N_{\boldsymbol{y}}(Q_{XY})^l\cdot N(Q_{XX'})^k\right]}{e^{n\cdot([R-I_Q(X;Y)]_++\epsilon)\cdot l}\cdot e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k}}, && (113)
\end{aligned}$$

where k and l are arbitrary positive integers, and where (108) follows from the definition of $\Xi(R,E_0,Q_{XX'})$ in (25). Step (111) is due to the fact that $\mathbb{P}\{X\geq a, Y\geq b\}\leq\mathbb{P}\{X\cdot Y\geq a\cdot b\}$, under the assumption that a,b are positive. In (112), we use the definition of $\beta(R,Q_Y)$ in (22), which implies that $\beta(R,Q_Y) \geq g(Q_{XY}) + [R - I_Q(X;Y)]_+$ and (113) follows from Markov's inequality. After optimizing over l and k,

$$\mathbb{P}\{\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y}) \cap \mathcal{G}_0\} \dot{\leq} \max_{Q_{XY}} \max_{Q_{XX'}} \inf_{l\in\mathbb{N}} \inf_{k\in\mathbb{N}} \frac{\mathbb{E}\left[N_{\boldsymbol{y}}(Q_{XY})^l \cdot N(Q_{XX'})^k\right]}{e^{n\cdot([R-I_Q(X;Y)]_+ + \epsilon)\cdot l} \cdot e^{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k}}. \tag{114}$$

For $S \geq 0$, a joint distribution Q_{UV}, and an integer $j \in \mathbb{N}$, define the following quantity

$$F(S, Q_{UV}, j) = \begin{cases} \exp\{nj\,(S - I_Q(U;V))\} & I_Q(U;V) < S \\ \exp\{n\,(S - I_Q(U;V))\} & I_Q(U;V) > S \end{cases}. \tag{115}$$

We use the following proposition, which is proved in Appendix G.

Proposition 4: Let $N(Q_{XX'})$ and $N_{\boldsymbol{y}}(Q_{XY})$ be as in (37) and (104), respectively. Then, for any $k,l \in \mathbb{N}$,

$$\mathbb{E}\left[N_{\boldsymbol{y}}(Q_{XY})^l N(Q_{XX'})^k\right] \dot{\leq} F(R,Q_{XY},l)\cdot F(2R,Q_{XX'},k). \tag{116}$$

Next, substituting the result of Proposition 4 back into (114) provides

$$\begin{aligned}\mathbb{P}\{\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y}) \cap \mathcal{G}_0\} &\dot{\leq} \max_{Q_{XY}} \inf_{l\in\mathbb{N}} \frac{e^{n\cdot(l\cdot[R-I_Q(X;Y)]_+ - [I_Q(X;Y)-R]_+)}}{\exp\{n\cdot([R-I_Q(X;Y)]_+ + \epsilon)\cdot l\}} \\ &\times \max_{Q_{XX'}} \inf_{k\in\mathbb{N}} \frac{e^{n\cdot(k\cdot[2R-I_Q(X;X')]_+ - [I_Q(X;X')-2R]_+)}}{\exp\{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k\}}. \end{aligned} \tag{117}$$

As for the left–hand term in (117), we have that

$$\begin{aligned} &-\frac{1}{n}\log \max_{Q_{XY}} \inf_{l\in\mathbb{N}} \frac{e^{n\cdot(l\cdot[R-I_Q(X;Y)]_+ - [I_Q(X;Y)-R]_+)}}{\exp\{n\cdot([R-I_Q(X;Y)]_+ + \epsilon)\cdot l\}} \\ &= -\frac{1}{n}\log \max_{Q_{XY}} \inf_{l\in\mathbb{N}} \exp\{-n\cdot([I_Q(X;Y)-R]_+ + l\epsilon)\} \quad (118) \\ &= \min_{Q_{XY}} \sup_{l\in\mathbb{N}} ([I_Q(X;Y)-R]_+ + l\epsilon) \quad (119) \\ &= \infty. \quad (120) \end{aligned}$$

For the right–hand term in (117), we get the following

$$\begin{aligned} &-\frac{1}{n}\log \max_{Q_{XX'}} \inf_{k\in\mathbb{N}} \frac{e^{n\cdot(k\cdot[2R-I_Q(X;X')]_+ - [I_Q(X;X')-2R]_+)}}{\exp\{n\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon)\cdot k\}} \\ &= \min_{Q_{XX'}} \sup_{k\in\mathbb{N}} (k\cdot(\Xi(R,E_0,Q_{XX'})+\epsilon \\ &\qquad - [2R - I_Q(X;X')]_+) + [I_Q(X;X')-2R]_+) \quad (121) \\ &= \min_{\substack{\{Q_{XX'}\in\mathcal{Q}(Q_X):\\ [2R-I_Q(X;X')]_+ \geq \Xi(R,E_0,Q_{XX'})+\epsilon\}}} [I_Q(X;X')-2R]_+ \quad (122) \\ &= E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon). \quad (123) \end{aligned}$$

Thus,

$$\mathbb{P}\{\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y}) \cap \mathcal{G}_0\} \dot{\leq} e^{-n\infty}\cdot \exp\{-n\cdot E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon)\}. \tag{124}$$

Final Steps: Finally, we continue from (97) and use the results of (103) and (124) to provide

$$\begin{aligned} &\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \leq E_0\right\} \\ &\dot{\geq} \mathbb{P}\{\mathcal{G}_0\} - \sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} \mathbb{P}\left\{\hat{\mathcal{B}}_n(\epsilon,m,m',\boldsymbol{y})\cap\mathcal{G}_0\right\} \quad (125) \\ &\dot{\geq} \exp\{-n\cdot E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon)\} \\ &\quad - \sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} e^{-n\infty}\cdot\exp\{-n\cdot E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon)\} \quad (126) \\ &\doteq \left(1 - e^{n2R}\cdot|\mathcal{Y}|^n\cdot e^{-n\infty}\right)\cdot \exp\{-n\cdot E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon)\} \quad (127) \\ &\doteq \exp\{-n\cdot E_{\text{lt}}^{\text{lb}}(R,E_0,\epsilon)\}. \quad (128) \end{aligned}$$

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

$$\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \leq E_0\right\} \dot{\geq} \exp\{-n\cdot E_{\text{lt}}^{\text{lb}}(R,E_0)\}, \tag{129}$$

which proves the lower bound of Theorem 1.

VI. Proof of the Upper Bound of Theorem 2

Let $Z_{mm'}(\boldsymbol{y})$, $\hat{\mathcal{B}}_n(\sigma)$, and $\hat{\mathcal{G}}_n(\sigma)$ be defined as in (74), (76), and (77), respectively. One of the main ingredients in the proof of the upper bound on the probability of the lower tail in Section V-A is the fact that $Z_m(\boldsymbol{y})$ is lower–bounded by $\exp\{n\alpha(R,\hat{P}_{\boldsymbol{y}})\}$ with a probability that approaches one double–exponentially fast. In order to prove an upper bound on the probability of the upper tail, we start by showing that $\exp\{n\beta(R,\hat{P}_{\boldsymbol{y}})\}$ serves as an upper bound on $Z_{mm'}(\boldsymbol{y})$, simultaneously for every $m \in \{0,1,\ldots,M-1\}$, $m' \in \{0,1,\ldots,M-1\}\setminus\{m\}$, and $\boldsymbol{y} \in \mathcal{Y}^n$, with probability that tends to one double–exponentially fast. More specifically, we have the following result, which is proved in Appendix H.

Proposition 5: For every $\sigma > 0$,

$$\mathbb{P}\left\{\hat{\mathcal{B}}_n(\sigma)\right\} \overset{\circ}{\leq} \exp\{-e^{n\sigma}\}. \tag{130}$$

We start with

$$\begin{aligned} &\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \geq E_0\right\} \\ &= \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \geq E_0\right\} \\ &\quad + \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{B}}_n(\sigma), -\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \geq E_0\right\} \quad (131) \\ &\leq \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n) \geq E_0\right\} \\ &\quad + \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{B}}_n(\sigma)\right\}. \quad (132) \end{aligned}$$

As for the first term,

$$\mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\right\}$$

$$= \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times \frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})} + e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})} + Z_{mm'}(\boldsymbol{y})} \leq e^{-n\cdot E_0}\right\} \quad (133)$$

$$\leq \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times \frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})} + e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})} + e^{n\cdot[\beta(R,\hat{P}_{\boldsymbol{y}})+\sigma]}} \leq e^{-n\cdot E_0}\right\} \quad (134)$$

$$\stackrel{\circ}{=} \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m) \times e^{n\cdot[\max\{g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}),\beta(R,\hat{P}_{\boldsymbol{y}})+\sigma\}-g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})]_+} \leq e^{-n\cdot E_0}\right\} \quad (135)$$

$$\stackrel{\circ}{=} \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m} e^{-n\cdot\tilde{\Lambda}(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R,\sigma)} \leq e^{-n\cdot E_0}\right\} \quad (136)$$

$$= \mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), \sum_{\mathcal{Q}(Q_X)} N(Q_{XX'})\cdot e^{-n\cdot\tilde{\Lambda}(Q_{XX'},R,\sigma)} \leq e^{n\cdot(R-E_0)}\right\} \quad (137)$$

$$\leq \mathbb{P}\left\{\sum_{\mathcal{Q}(Q_X)} N(Q_{XX'})\cdot e^{-n\cdot\tilde{\Lambda}(Q_{XX'},R,\sigma)} \leq e^{n\cdot(R-E_0)}\right\}, \quad (138)$$

where (133) follows from the definitions of the probability of error and $Z_{mm'}(\boldsymbol{y})$ in (11) and (74), respectively. In (134), the definition of the set $\hat{\mathcal{G}}_n(\sigma)$ in (77) was used, in (135), the exponential equivalence $e^{nB}/(e^{nA}+e^{nB}+e^{nC}) \doteq \exp\{-n\cdot[\max\{A,C\}-B]_+\}$, in (136), the method of types and the definition of $\tilde{\Lambda}(Q_{XX'},R,\sigma)$ in (78), in (137), the definition of the type class enumerators $N(Q_{XX'})$ in (37), and in (138), the event $\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma)$ was taken out.

Next,

$$\mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\right\}$$

$$\stackrel{\circ}{\leq} \mathbb{P}\left\{\sum_{\mathcal{Q}(Q_X)} N(Q_{XX'})\cdot e^{-n\cdot\tilde{\Lambda}(Q_{XX'},R,\sigma)} \leq e^{n\cdot(R-E_0)}\right\} \quad (139)$$

$$\stackrel{\circ}{=} \mathbb{P}\left\{\max_{\mathcal{Q}(Q_X)} N(Q_{XX'})\cdot e^{-n\cdot\tilde{\Lambda}(Q_{XX'},R,\sigma)} \leq e^{n\cdot(R-E_0)}\right\} \quad (140)$$

$$= \mathbb{P}\left\{\bigcap_{\mathcal{Q}(Q_X)} \left\{N(Q_{XX'}) \leq e^{n\cdot(\tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0)}\right\}\right\}, \quad (141)$$

where (140) is due to the SME.

If E_0 is relatively small, then for every $Q_{XX'} \in \mathcal{Q}(Q_X)$, either $I_Q(X;X') \geq 2R$ or $2R - I_Q(X;X') \leq \tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0$, and we have an intersection of polynomially many events whose probabilities all tend to one. Hence, for every $\sigma > 0$, we assume that E_0 is sufficiently large, so there must exist at least one $Q_{XX'} \in \mathcal{Q}(Q_X)$ for which $I_Q(X;X') \leq 2R$ and $\tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0 \leq 2R - I_Q(X;X')$, such that (141) decays double exponentially fast, according to Lemma 2 in Appendix B. We define the set

$$\tilde{\mathcal{V}}(R,E_0,\sigma) \stackrel{\Delta}{=} \{Q_{XX'} \in \mathcal{Q}(Q_X): \; I_Q(X;X') \leq 2R, \; \tilde{\Lambda}(Q_{XX'},R,\sigma) + I_Q(X;X') - R \leq E_0\}. \quad (142)$$

Then,

$$\mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\right\}$$

$$\stackrel{\circ}{\leq} \mathbb{P}\left\{\bigcap_{\mathcal{Q}(Q_X)} \left\{N(Q_{XX'}) \leq e^{n\cdot(\tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0)}\right\}\right\} \quad (143)$$

$$\leq \mathbb{P}\left\{\bigcap_{\tilde{\mathcal{V}}(R,E_0,\sigma)} \left\{N(Q_{XX'}) \leq e^{n\cdot(\tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0)}\right\}\right\}. \quad (144)$$

Since $\tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0 \leq 2R - I_Q(X;X')$, we obtain

$$\mathbb{P}\left\{\bigcap_{\tilde{\mathcal{V}}(R,E_0,\sigma)} \left\{N(Q_{XX'}) \leq e^{n\cdot(\tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0)}\right\}\right\} \quad (145)$$

$$\leq \min_{\tilde{\mathcal{V}}(R,E_0,\sigma)} \mathbb{P}\left\{N(Q_{XX'}) \leq e^{n\cdot(\tilde{\Lambda}(Q_{XX'},R,\sigma)+R-E_0)}\right\} \quad (146)$$

$$\stackrel{\circ}{\leq} \min_{\tilde{\mathcal{V}}(R,E_0,\sigma)} \exp\left\{-\min\left(e^{n(2R-I_Q(X;X'))}, e^{nR}\right)\right\} \quad (147)$$

$$= \min_{\tilde{\mathcal{V}}(R,E_0,\sigma)} \exp\left\{-e^{n\cdot\min\{2R-I_Q(X;X'),R\}}\right\} \quad (148)$$

$$= \exp\left\{-e^{n\cdot\max_{\tilde{\mathcal{V}}(R,E_0,\sigma)}\min\{2R-I_Q(X;X'),R\}}\right\}, \quad (149)$$

where (147) follows from Lemma 2 in Appendix B. Let us define

$$E_1(R,E_0,\sigma) = \max_{Q_{XX'}\in\tilde{\mathcal{V}}(R,E_0,\sigma)} \min\{2R - I_Q(X;X'), R\}, \quad (150)$$

such that

$$\mathbb{P}\left\{\mathcal{C}_n \in \hat{\mathcal{G}}_n(\sigma), -\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0\right\} \stackrel{\circ}{\leq} \exp\left\{-\exp\left\{n\cdot E_1(R,E_0,\sigma)\right\}\right\}. \quad (151)$$

Final Steps: Finally, it follows from (151) and Proposition 5 that

$$\begin{aligned}\mathbb{P}&\left\{-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)\geq E_0\right\}\\&\leq \mathbb{P}\left\{\mathcal{C}_n\in\hat{\mathcal{G}}_n(\sigma),-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)\geq E_0\right\}\\&\quad+\mathbb{P}\left\{\mathcal{C}_n\in\hat{\mathcal{B}}_n(\sigma)\right\} \qquad (152)\\&\overset{\circ}{\leq}\exp\left\{-e^{n\cdot E_1(R,E_0,\sigma)}\right\}+\exp\left\{-e^{n\sigma}\right\} \qquad (153)\\&\overset{\circ}{=}\exp\left\{-\exp\{n\cdot\min[E_1(R,E_0,\sigma),\sigma]\}\right\}. \qquad (154)\end{aligned}$$

As a last step, we optimize over $\sigma>0$, which resulting in

$$\begin{aligned}\mathbb{P}&\left\{-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)> E_0\right\}\\&\overset{\circ}{\leq}\exp\left\{-\exp\left\{n\cdot\sup_{\sigma>0}\min[E_1(R,E_0,\sigma),\sigma]\right\}\right\}. \qquad (155)\end{aligned}$$

A Simplified Expression: Note that $E_1(R,E_0,\sigma)$ is continuous and monotonically non–increasing in σ, hence we can solve for the optimal $\sigma>0$ by finding the maximal σ for which $\sigma\leq E_1(R,E_0,\sigma)$. Let us abbreviate $I_Q(X;X')$ by I_Q, and then

$$\begin{aligned}&E_1(R,E_0,\sigma)\\&=\max_{Q_{XX'}\in\tilde{\mathcal{V}}(R,E_0,\sigma)}\min\{2R-I_Q,R\} \qquad (156)\\&=\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R\}}\inf_{\mu\geq 0}\Big\{\min\{2R-I_Q,R\}\\&\qquad+\mu\cdot(E_0-\tilde{\Lambda}(Q_{XX'},R,\sigma)-I_Q+R)\Big\} \qquad (157)\\&=\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R\}}\inf_{\mu\geq 0}\{\min\{2R-I_Q,R\}\\&\qquad+\mu\cdot(E_0-\Lambda(Q_{XX'},R)-\sigma-I_Q+R)\} \qquad (158)\\&=\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R\}}\inf_{\mu\geq 0}\{\min\{2R-I_Q,R\}\\&\qquad+\mu\cdot(E_0-\Lambda(Q_{XX'},R)-I_Q+R)-\mu\sigma\}, \qquad (159)\end{aligned}$$

where (157) is due to (142) and the fact that $\max_{\{Q:\ g(Q)\geq 0\}}f(Q)=\max_Q\inf_{\mu\geq 0}\{f(Q)+\mu\cdot g(Q)\}$ and (158) is true thanks to (91). Now, we would like to solve for

$$\begin{aligned}\sigma\leq&\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R\}}\inf_{\mu\geq 0}\{\min\{2R-I_Q,R\}\\&+\mu\cdot(E_0-\Lambda(Q_{XX'},R)-I_Q+R)-\mu\sigma\}, \qquad (160)\end{aligned}$$

which is equivalent to the statement

$$\begin{aligned}&\exists Q_{XX'}\in\mathcal{Q}(Q_X)\text{ s.t. }I_Q\leq 2R,\quad\forall\mu\geq 0:\\&\sigma\leq\min\{2R-I_Q,R\}\\&\qquad+\mu\cdot(E_0-\Lambda(Q_{XX'},R)-I_Q+R)-\mu\sigma, \qquad (161)\end{aligned}$$

or,

$$\begin{aligned}&\exists Q_{XX'}\in\mathcal{Q}(Q_X)\text{ s.t. }I_Q\leq 2R,\quad\forall\mu\geq 0:\\&\sigma\leq\frac{1}{1+\mu}\cdot[\min\{2R-I_Q,R\}\\&\qquad+\mu\cdot(E_0-\Lambda(Q_{XX'},R)-I_Q+R)], \qquad (162)\end{aligned}$$

or, equivalently,

$$\begin{aligned}\sigma\leq&\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R\}}\inf_{\mu\geq 0}\frac{1}{1+\mu}\cdot[\min\{2R-I_Q,R\}\\&+\mu\cdot(E_0-\Lambda(Q_{XX'},R)-I_Q+R)]. \qquad (163)\end{aligned}$$

For simplicity, let us denote

$$A=\min\{2R-I_Q,R\}, \qquad (164)$$

$$B=E_0-\Lambda(Q_{XX'},R)-I_Q+R, \qquad (165)$$

such that

$$\begin{aligned}\sigma&\leq\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R\}}\inf_{\mu\geq 0}\left\{\frac{A+\mu B}{1+\mu}\right\} \qquad (166)\\&=\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R\}}\min\{A,B\} \qquad (167)\\&=\max\left\{\begin{array}{l}\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R,\ B\geq 0\}}\min\{A,B\}\\\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R,\ B<0\}}\min\{A,B\}\end{array}\right. \qquad (168)\\&=\max\left\{\begin{array}{l}\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R,\ B\geq 0\}}\min\{A,B\}\\\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R,\ B<0\}}B\end{array}\right. \qquad (169)\\&=\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q\leq 2R,\ B\geq 0\}}\min\{A,B\} \qquad (170)\\&=\max_{Q_{XX'}\in\mathcal{V}(R,E_0)}\min\{2R-I_Q,\\&\qquad E_0-\Lambda(Q_{XX'},R)-I_Q+R,R\} \qquad (171)\\&=E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0), \qquad (172)\end{aligned}$$

where (169) and (170) are due to the fact that $A\geq 0$, while (171) and (172) follow from the definitions in (42) and (44), respectively. Thus,

$$\begin{aligned}\mathbb{P}&\left\{-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)\geq E_0\right\}\\&\overset{\circ}{\leq}\exp\left\{-\exp\left\{n\cdot\sup_{\sigma>0}\min[E_1(R,E_0,\sigma),\sigma]\right\}\right\} \qquad (173)\\&=\exp\left\{-\exp\left\{n\cdot\sup_{0<\sigma\leq E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0)}\sigma\right\}\right\} \qquad (174)\\&=\exp\left\{-e^{n\cdot E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0)}\right\}, \qquad (175)\end{aligned}$$

and the proof of the upper bound of Theorem 2 is complete.

VII. Proof of the Lower Bound of Theorem 2

Let the sets $\mathcal{B}_\epsilon(m,\boldsymbol{y})$ and $\mathcal{B}_\epsilon$ be as defined in (52) and (54), respectively. Also define $\mathcal{G}_\epsilon(m,\boldsymbol{y})=\mathcal{B}_\epsilon^{\mathrm{c}}(m,\boldsymbol{y})$ and $\mathcal{G}_\epsilon=\mathcal{B}_\epsilon^{\mathrm{c}}$. Let $E_0>0$ be given. Then,

$$\begin{aligned}\mathbb{P}&\left\{-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n)\geq E_0\right\}\\&=\mathbb{P}\Bigg\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}W(\boldsymbol{y}|\boldsymbol{x}_m)\\&\qquad\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+Z_m(\boldsymbol{y})}\leq e^{-n\cdot E_0}\Bigg\} \qquad (176)\\&\geq\mathbb{P}\Bigg\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}W(\boldsymbol{y}|\boldsymbol{x}_m)\\&\qquad\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+Z_m(\boldsymbol{y})}\leq e^{-n\cdot E_0},\ \mathcal{C}_n\in\mathcal{G}_\epsilon\Bigg\} \qquad (177)\end{aligned}$$

$$\stackrel{\circ}{\geq} \mathbb{P}\left\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\exp\{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}}, R-\epsilon)\} \leq e^{-n\cdot E_0},\ \mathcal{C}_n \in \mathcal{G}_\epsilon\right\}, \tag{178}$$

where (176) follows from the definitions of the probability of error and $Z_m(\boldsymbol{y})$ in (11) and (51), respectively. Step (178) follows from the same considerations as in eqs. (59)–(62). Now, define the event

$$\mathcal{E}_0 = \left\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\exp\{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}}, R-\epsilon)\} \leq e^{-n\cdot E_0}\right\}, \tag{179}$$

such that, continuing from (178),

$$\begin{aligned}
&\mathbb{P}\{\mathcal{C}_n \in \mathcal{E}_0,\ \mathcal{C}_n \in \mathcal{G}_\epsilon\} \\
&= \mathbb{P}\left\{\bigcap_{\bar{m}=0}^{M-1}\bigcap_{\boldsymbol{y}\in\mathcal{Y}^n}\mathcal{G}_\epsilon(\bar{m},\boldsymbol{y})\middle|\mathcal{E}_0\right\}\cdot\mathbb{P}\{\mathcal{E}_0\} &(180)\\
&= \left(1-\mathbb{P}\left\{\bigcup_{\bar{m}=0}^{M-1}\bigcup_{\boldsymbol{y}\in\mathcal{Y}^n}\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})\middle|\mathcal{E}_0\right\}\right)\cdot\mathbb{P}\{\mathcal{E}_0\} &(181)\\
&\geq \left(1-\sum_{\bar{m}=0}^{M-1}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\{\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})|\mathcal{E}_0\}\right)\cdot\mathbb{P}\{\mathcal{E}_0\} &(182)\\
&= \mathbb{P}\{\mathcal{E}_0\} - \sum_{\bar{m}=0}^{M-1}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\{\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})\cap\mathcal{E}_0\}. &(183)
\end{aligned}$$

Lower–bounding $\mathbb{P}\{\mathcal{E}_0\}$ in (183)*:* First of all, note that

$$\begin{aligned}
&\mathbb{P}\{\mathcal{E}_0\} \\
&= \mathbb{P}\left\{\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}e^{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R-\epsilon)} \leq e^{-n\cdot E_0}\right\} &(184)\\
&= \mathbb{P}\left\{\sum_{\mathcal{Q}(Q_X)}N(Q_{XX'})e^{-n\Gamma(Q_{XX'},R-\epsilon)} \leq e^{n\cdot(R-E_0)}\right\} &(185)\\
&\stackrel{\circ}{=} \mathbb{P}\left\{\max_{\mathcal{Q}(Q_X)}N(Q_{XX'})e^{-n\Gamma(Q_{XX'},R-\epsilon)} \leq e^{n\cdot(R-E_0)}\right\} &(186)\\
&= \mathbb{P}\left\{\bigcap_{\mathcal{Q}(Q_X)}\left\{N(Q_{XX'}) \leq e^{n\cdot(\Gamma(Q_{XX'},R-\epsilon)+R-E_0)}\right\}\right\}, &(187)
\end{aligned}$$

where in (185), the definition of $N(Q_{XX'})$ in (37) was used, and (186) is due to the SME in (7).

Now, if there exists at least one $Q_{XX'} \in \mathcal{Q}(Q_X)$ for which $I_Q(X;X') < 2R$ and $2R - I_Q(X;X') > \Gamma(Q_{XX'}, R-\epsilon) + R - E_0$, then this $Q_{XX'}$ alone is responsible for a double exponential decay of the probability of the event $\{N(Q_{XX'}) \leq e^{n\cdot(\Gamma(Q_{XX'},R-\epsilon)+R-E_0)}\}$ (thanks to Lemma 2 in Appendix B), such that the probability in (187), which is of the intersection over all $Q_{XX'} \in \mathcal{Q}(Q_X)$, decays double exponentially fast. On the other hand, if for every $Q_{XX'} \in \mathcal{Q}(Q_X)$, either $I_Q(X;X') \geq 2R$ or $2R - I_Q(X;X') \leq \Gamma(Q_{XX'}, R-\epsilon) + R - E_0$, then we have an intersection of polynomially many events whose probabilities all tend to one. Thus, this probability is exponentially equal to one if and only if for every $Q_{XX'} \in \mathcal{Q}(Q_X)$, either $I_Q(X;X') \geq 2R$ or $2R - I_Q(X;X') \leq \Gamma(Q_{XX'}, R-\epsilon) + R - E_0$, or equivalently,

$$2R \leq \min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\{I_Q(X;X') + [\Gamma(Q_{XX'}, R-\epsilon) + R - E_0]_+\}. \tag{188}$$

Let us now find what is the maximum value of E_0 for which this inequality holds true. The condition is equivalent to

$$\min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\max_{0\leq a\leq 1}\{I_Q(X;X') + a(\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)\} \geq 2R, \tag{189}$$

or

$$\forall Q_{XX'} \in \mathcal{Q}(Q_X)\ \exists a \in [0,1]: \quad I_Q(X;X') + a(\Gamma(Q_{XX'}, R-\epsilon) + R - E_0) \geq 2R, \tag{190}$$

or

$$\forall Q_{XX'} \in \mathcal{Q}(Q_X)\ \exists a \in [0,1]: \quad \Gamma(Q_{XX'}, R-\epsilon) + R + \frac{1}{a}(I_Q(X;X') - 2R) \geq E_0, \tag{191}$$

or, equivalently,

$$\begin{aligned}
&E_0 \\
&\leq \min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\max_{0\leq a\leq 1}\left\{\Gamma(Q_{XX'}, R-\epsilon) + R + \frac{1}{a}(I_Q(X;X') - 2R)\right\} &(192)\\
&= \min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\left[\Gamma(Q_{XX'}, R-\epsilon) + R + \begin{cases} I_Q(X;X') - 2R & 2R \geq I_Q(X;X') \\ \infty & 2R < I_Q(X;X')\end{cases}\right] &(193)\\
&= \min_{\left\{\substack{Q_{XX'}\in\mathcal{Q}(Q_X):\\ I_Q(X;X')\leq 2R}\right\}}\{\Gamma(Q_{XX'}, R-\epsilon) + I_Q(X;X') - R\} &(194)\\
&\leq E_{\text{trc}}(R). &(195)
\end{aligned}$$

Thus, we assume that $E_0 > E_{\text{trc}}(R)$, which ensures that there exists at least one $Q_{XX'} \in \mathcal{Q}(Q_X)$ for which $I_Q(X;X') \leq 2R$ and $\Gamma(Q_{XX'}, R-\epsilon) + R - E_0 \leq 2R - I_Q(X;X')$, such that the probability in (187) decays double exponentially fast. Define

$$\begin{aligned}
\mathcal{A}_1 &= \{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q(X;X') > 2R\} &(196)\\
\mathcal{A}_2 &= \{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q(X;X') \leq 2R,\\
&\qquad \Gamma(Q_{XX'}, R-\epsilon) + I_Q(X;X') - R \leq E_0 + \epsilon\} &(197)\\
\mathcal{A}_3 &= \{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q(X;X') \leq 2R,\\
&\qquad \Gamma(Q_{XX'}, R-\epsilon) + I_Q(X;X') - R > E_0 + \epsilon\}. &(198)
\end{aligned}$$

Defining the events

$$\mathcal{F}_0 = \bigcap_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \{N(Q_{XX'}) = 0\}, \tag{199}$$

and,

$$\mathcal{F}(Q_{XX'}) = \left\{N(Q_{XX'}) \le e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)}\right\}, \tag{200}$$

then considering the probability in (187), we have that

$$\begin{aligned}
&\mathbb{P}\left\{\bigcap_{Q_{XX'} \in \mathcal{Q}(Q_X)} \mathcal{F}(Q_{XX'})\right\} \\
&= \mathbb{P}\left\{\bigcap_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3} \mathcal{F}(Q_{XX'})\right\} && (201) \\
&\ge \mathbb{P}\left\{\bigcap_{Q_{XX'} \in \mathcal{A}_3} \mathcal{F}(Q_{XX'}), \bigcap_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \{N(Q_{XX'}) = 0\}\right\} && (202) \\
&= \mathbb{P}\left\{\bigcap_{Q_{XX'} \in \mathcal{A}_3} \mathcal{F}(Q_{XX'}) \middle| \mathcal{F}_0\right\} \cdot \mathbb{P}\{\mathcal{F}_0\} && (203) \\
&= \left(1 - \mathbb{P}\left\{\bigcup_{Q_{XX'} \in \mathcal{A}_3} \mathcal{F}^c(Q_{XX'}) \middle| \mathcal{F}_0\right\}\right) \cdot \mathbb{P}\{\mathcal{F}_0\} && (204) \\
&\ge \left(1 - \sum_{Q_{XX'} \in \mathcal{A}_3} \mathbb{P}\{\mathcal{F}^c(Q_{XX'}) | \mathcal{F}_0\}\right) \cdot \mathbb{P}\{\mathcal{F}_0\}. && (205)
\end{aligned}$$

Next, it follows from Markov's inequality that

$$\begin{aligned}
&\mathbb{P}\left\{N(Q_{XX'}) \ge e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)} \middle| \mathcal{F}_0\right\} && (206) \\
&\le \frac{\mathbb{E}[N(Q_{XX'}) | \mathcal{F}_0]}{e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)}} && (207) \\
&= \frac{\mathbb{E}\left[\sum_{m=0}^{M-1} \sum_{m' \ne m} \mathcal{I}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'})\} \middle| \mathcal{F}_0\right]}{e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)}} && (208) \\
&= \frac{\sum_{m=0}^{M-1} \sum_{m' \ne m} \mathbb{P}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'}) | \mathcal{F}_0\}}{e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)}} && (209) \\
&\le \frac{e^{n2R} \cdot \mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'}) | \mathcal{F}_0\}}{e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)}}. && (210)
\end{aligned}$$

We continue from (205) and get that

$$\begin{aligned}
&\mathbb{P}\left\{\bigcap_{Q_{XX'} \in \mathcal{Q}(Q_X)} \mathcal{F}(Q_{XX'})\right\} \\
&\ge \left(1 - \sum_{Q_{XX'} \in \mathcal{A}_3} \frac{e^{n2R} \cdot \mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'}) | \mathcal{F}_0\}}{e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)}}\right) \\
&\qquad \cdot \mathbb{P}\{\mathcal{F}_0\} && (211) \\
&= \mathbb{P}\{\mathcal{F}_0\} \\
&\quad - \sum_{Q_{XX'} \in \mathcal{A}_3} \frac{e^{n2R} \cdot \mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'}), \mathcal{F}_0\}}{e^{n \cdot (\Gamma(Q_{XX'}, R-\epsilon) + R - E_0)}}. && (212)
\end{aligned}$$

In order to upper–bound the probabilities in the summation in (212), we define the following truncated enumerators

$$\tilde{N}(Q_{XX'}) \triangleq \sum_{m=2}^{M-1} \sum_{m' \in \{2,3,\ldots,M-1\} \setminus \{m\}} \mathcal{I}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\}, \tag{213}$$

and the event

$$\mathcal{F}_1 = \bigcap_{Q_{XX'} \in \mathcal{A}_2} \left\{\tilde{N}(Q_{XX'}) = 0\right\}. \tag{214}$$

Then,

$$\begin{aligned}
&\mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'}), \mathcal{F}_0\} \\
&= \mathbb{P}\Bigg\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'}), \\
&\qquad \bigcap_{\hat{Q}_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \left\{N(\hat{Q}_{XX'}) = 0\right\}\Bigg\} && (215) \\
&= \mathbb{P}\Bigg\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'}), \bigcap_{\hat{Q}_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \bigcap_{m=0}^{M-1} \\
&\qquad \bigcap_{m' \in \{0,1,\ldots,M-1\} \setminus \{m\}} \left\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \notin \mathcal{T}(\hat{Q}_{XX'})\right\}\Bigg\} && (216) \\
&\le \mathbb{P}\Bigg\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'}), \bigcap_{\hat{Q}_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \bigcap_{m=2}^{M-1} \\
&\qquad \bigcap_{m' \in \{2,3,\ldots,M-1\} \setminus \{m\}} \left\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \notin \mathcal{T}(\hat{Q}_{XX'})\right\}\Bigg\} && (217) \\
&= \mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'})\} \\
&\quad \times \mathbb{P}\Bigg\{\bigcap_{\hat{Q}_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \bigcap_{m=2}^{M-1} \\
&\qquad \bigcap_{m' \in \{2,3,\ldots,M-1\} \setminus \{m\}} \left\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \notin \mathcal{T}(\hat{Q}_{XX'})\right\}\Bigg\} && (218) \\
&= \mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'})\} \\
&\quad \cdot \mathbb{P}\left\{\bigcap_{\hat{Q}_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \left\{\tilde{N}(\hat{Q}_{XX'}) = 0\right\}\right\} && (219) \\
&\le \mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'})\} \\
&\quad \cdot \mathbb{P}\left\{\bigcap_{\hat{Q}_{XX'} \in \mathcal{A}_2} \left\{\tilde{N}(\hat{Q}_{XX'}) = 0\right\}\right\} && (220) \\
&= \mathbb{P}\{(\boldsymbol{X}_0, \boldsymbol{X}_1) \in \mathcal{T}(Q_{XX'})\} \cdot \mathbb{P}\{\mathcal{F}_1\}. && (221)
\end{aligned}$$

Substituting it back into (212), now yields

$$\mathbb{P}\left\{\bigcap_{Q_{XX'}\in\mathcal{Q}(Q_X)}\mathcal{F}(Q_{XX'})\right\}$$
$$\geq \mathbb{P}\{\mathcal{F}_0\} - \sum_{Q_{XX'}\in\mathcal{A}_3}\frac{e^{n2R}\cdot\mathbb{P}\{(\boldsymbol{X}_0,\boldsymbol{X}_1)\in\mathcal{T}(Q_{XX'}),\mathcal{F}_0\}}{e^{n\cdot(\Gamma(Q_{XX'},R-\epsilon)+R-E_0)}} \tag{222}$$
$$\geq \mathbb{P}\{\mathcal{F}_0\} - \sum_{Q_{XX'}\in\mathcal{A}_3}\frac{e^{n2R}\cdot\mathbb{P}\{(\boldsymbol{X}_0,\boldsymbol{X}_1)\in\mathcal{T}(Q_{XX'})\}\cdot\mathbb{P}\{\mathcal{F}_1\}}{e^{n\cdot(\Gamma(Q_{XX'},R-\epsilon)+R-E_0)}} \tag{223}$$
$$= \mathbb{P}\{\mathcal{F}_0\} - \mathbb{P}\{\mathcal{F}_1\}\cdot\sum_{Q_{XX'}\in\mathcal{A}_3}\frac{e^{n2R}\cdot\mathbb{P}\{(\boldsymbol{X}_0,\boldsymbol{X}_1)\in\mathcal{T}(Q_{XX'})\}}{e^{n\cdot(\Gamma(Q_{XX'},R-\epsilon)+R-E_0)}}. \tag{224}$$

Generally, it follows that $\mathbb{P}\{\mathcal{F}_0\} \leq \mathbb{P}\{\mathcal{F}_1\}$. First, we lower–bound $\mathbb{P}\{\mathcal{F}_0\}$. The following proposition is proved in Appendix I:

Proposition 6: If $E_0 < E_{\text{ex}}(R)$, then

$$\mathbb{P}\{\mathcal{F}_0\} \overset{\circ}{\geq} \exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\}. \tag{225}$$

In addition, we can easily prove that under the condition of $E_0 < E_{\text{ex}}(R)$, $\mathbb{P}\{\mathcal{F}_1\}$ can be upper–bounded by the same expression that lower–bounds $\mathbb{P}\{\mathcal{F}_0\}$. We have that

$$\mathbb{P}\{\mathcal{F}_1\}$$
$$= \mathbb{P}\left\{\bigcap_{Q_{XX'}\in\mathcal{A}_2}\left\{\tilde{N}(Q_{XX'})=0\right\}\right\} \tag{226}$$
$$\leq \min_{Q_{XX'}\in\mathcal{A}_2}\mathbb{P}\left\{\tilde{N}(Q_{XX'})=0\right\} \tag{227}$$
$$\overset{\circ}{\leq} \min_{Q_{XX'}\in\mathcal{A}_2}\exp\left\{-\min\left(e^{n(2R-I_Q(X;X'))},e^{nR}\right)\right\} \tag{228}$$
$$= \min_{Q_{XX'}\in\mathcal{A}_2}\exp\left\{-e^{n(2R-I_Q(X;X'))}\right\} \tag{229}$$
$$= \exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\}, \tag{230}$$

where (228) is due to Lemma 2 in Appendix B and (229) follows from the fact that $E_0 < E_{\text{ex}}(R)$ is equivalent to $\min_{Q_{XX'}\in\mathcal{A}_2} I_Q(X;X') > R$ (Appendix I). Hence,

$$\mathbb{P}\{\mathcal{F}_0\} \doteq \mathbb{P}\{\mathcal{F}_1\} \doteq \exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\}. \tag{231}$$

Using the definition of the set $\mathcal{A}_3$ provides

$$\mathbb{P}\{\mathcal{E}_0\}$$
$$\doteq \mathbb{P}\left\{\bigcap_{Q_{XX'}\in\mathcal{Q}(Q_X)}\mathcal{F}(Q_{XX'})\right\} \tag{232}$$
$$\geq \mathbb{P}\{\mathcal{F}_0\} - \mathbb{P}\{\mathcal{F}_1\}\cdot\sum_{Q_{XX'}\in\mathcal{A}_3}\frac{e^{n2R}\cdot\mathbb{P}\{(\boldsymbol{X}_0,\boldsymbol{X}_1)\in\mathcal{T}(Q_{XX'})\}}{e^{n\cdot(\Gamma(Q_{XX'},R-\epsilon)+R-E_0)}} \tag{233}$$
$$\doteq \left(1-\sum_{Q_{XX'}\in\mathcal{A}_3}\frac{e^{n\cdot(2R-I_Q(X;X'))}}{e^{n\cdot(\Gamma(Q_{XX'},R-\epsilon)+R-E_0)}}\right)\cdot\exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\} \tag{234}$$
$$\doteq \left(1-e^{-n\epsilon}\right)\cdot\exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\} \tag{235}$$
$$\doteq \exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\}. \tag{236}$$

Upper–bounding $\mathbb{P}\{\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})\cap\mathcal{E}_0\}$ in (183)*:* Recall that

$$\mathbb{P}\{\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})\cap\mathcal{E}_0\}$$
$$= \mathbb{P}\left\{\sum_{\tilde{m}\in\{0,1,\ldots,M-1\}\setminus\{\bar{m}\}} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}}\boldsymbol{y}})} \leq e^{n\cdot\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})}, \sum_{m=0}^{M-1}\sum_{m'\neq m} e^{-n\Gamma(\hat{P}_{\boldsymbol{X}_m\boldsymbol{X}_{m'}},R-\epsilon)} \leq e^{n\cdot(R-E_0)}\right\}. \tag{237}$$

In order to upper–bound this probability, we do the following. In the first event, instead of summing over $\{0,1,\ldots,M-1\}\setminus\{\bar{m}\}$, we sum over $\{\lfloor M/2\rfloor,\lfloor M/2\rfloor+1,\ldots,M-1\}\setminus\{\bar{m}\}$, and in the second event, instead of summing over $\{(m,m') :\ m,m'\in\{0,1,\ldots,M-1\},\ m\neq m'\}$, we sum over $\mathcal{N}^2 = \{(m,m') :\ m,m'\in\{0,1,\ldots,\lfloor M/2\rfloor-1\}, m\neq m'\}$, hence, the two events become independent:

$$\mathbb{P}\{\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})\cap\mathcal{E}_0\}$$
$$\leq \mathbb{P}\left\{\sum_{\tilde{m}\in\{\lfloor M/2\rfloor,\ldots,M-1\}\setminus\{\bar{m}\}} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}}\boldsymbol{y}})} \leq e^{n\cdot\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})}\right\} \times\mathbb{P}\left\{\sum_{(m,m')\in\mathcal{N}^2} e^{-n\Gamma(\hat{P}_{\boldsymbol{X}_m\boldsymbol{X}_{m'}},R-\epsilon)} \leq e^{n\cdot(R-E_0)}\right\}. \tag{238}$$

As for the first factor in (238), note that its sum has exponentially many terms as $Z_m(\boldsymbol{y})$, and hence is also upper–bounded as in (53). The second factor in (238) can be upper–bounded using similar analysis as in the proof in Section VI, which results an upper bound similar to (230). Thus,

$$\begin{aligned}&\mathbb{P}\left\{\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})\cap\mathcal{E}_0\right\}\\&\leq \exp\{-e^{n\epsilon}+n\epsilon+1\}\\&\quad\cdot\exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\}.\end{aligned}\tag{239}$$

Final Steps: Finally, we continue from (183) and use the results of (234) and (239) to obtain

$$\begin{aligned}&\mathbb{P}\left\{-\frac{1}{n}\log P_e(\mathcal{C}_n)\geq E_0\right\}\\&\overset{\circ}{\geq}\mathbb{P}\left\{\mathcal{E}_0\right\}-\sum_{\bar{m}=0}^{M-1}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\left\{\mathcal{B}_\epsilon(\bar{m},\boldsymbol{y})\cap\mathcal{E}_0\right\} &(240)\\&\overset{\circ}{\geq}\exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\}\\&\quad-\sum_{\bar{m}=0}^{M-1}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\exp\{-e^{n\epsilon}+n\epsilon+1\}\\&\quad\cdot\exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\} &(241)\\&=\left(1-e^{nR}\cdot|\mathcal{Y}|^n\cdot\exp\{-e^{n\epsilon}+n\epsilon+1\}\right)\\&\quad\cdot\exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\} &(242)\\&\overset{\circ}{=}\exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{A}_2}\{2R-I_Q(X;X')\}\right\}\right\}, &(243)\end{aligned}$$

which proves the lower bound of Theorem 2.

Appendix A

Preliminaries

The main purpose of this appendix is to provide the general setting and the main results that are borrowed from [19].

Let $\{U_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathcal{K}}$, where $\mathcal{K}$ is a set of multidimensional indexes, be a family of Bernoulli random variables. Let G be a dependency graph for $\{U_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathcal{K}}$, i.e., a graph with vertex set $\mathcal{K}$ such that if $\mathcal{A}$ and $\mathcal{B}$ are two disjoint subsets of $\mathcal{K}$, and G contains no edge between $\mathcal{A}$ and $\mathcal{B}$, then the families $\{U_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathcal{A}}$ and $\{U_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathcal{B}}$ are independent. Let $S=\sum_{\boldsymbol{k}\in\mathcal{K}}U_{\boldsymbol{k}}$ and $\Delta=\mathbb{E}[S]$. Moreover, we write $\boldsymbol{i}\sim\boldsymbol{j}$ if $(\boldsymbol{i},\boldsymbol{j})$ is an edge in the dependency graph G. Let

$$\Phi=\max_{\boldsymbol{i}\in\mathcal{K}}\mathbb{E}[U_{\boldsymbol{i}}],\tag{A.1}$$

$$\Omega_{\boldsymbol{i}}=\sum_{\boldsymbol{j}\in\mathcal{K},\boldsymbol{j}\sim\boldsymbol{i}}\mathbb{E}[U_{\boldsymbol{j}}],\tag{A.2}$$

$$\Omega=\max_{\boldsymbol{i}\in\mathcal{K}}\sum_{\boldsymbol{j}\in\mathcal{K},\boldsymbol{j}\sim\boldsymbol{i}}\mathbb{E}[U_{\boldsymbol{j}}],\tag{A.3}$$

and

$$\Theta=\frac{1}{2}\sum_{\boldsymbol{i}\in\mathcal{K}}\sum_{\boldsymbol{j}\in\mathcal{K},\boldsymbol{j}\sim\boldsymbol{i}}\mathbb{E}[U_{\boldsymbol{i}}U_{\boldsymbol{j}}].\tag{A.4}$$

The following result will be used in the proof of Lemma 2 in Appendix B:

Fact 1: With notations as above, [19, Th. 10] states that for any $0\leq a\leq 1$,

$$\begin{aligned}&\mathbb{P}\{S\leq a\Delta\}\\&\quad\leq\exp\left\{-\min\left((1-a)^2\frac{\Delta^2}{8\Theta+2\Delta},(1-a)\frac{\Delta}{6\Omega}\right)\right\}.\end{aligned}\tag{A.5}$$

The following result will be used in the proof of Lemma 6 in Appendix B:

Fact 2: With notations as above, [19, Th. 3] states that,

$$\mathbb{P}\{S=0\}\leq\exp\left\{-\min\left(\frac{\Delta^2}{8\Theta},\frac{\Delta}{6\Omega},\frac{\Delta}{2}\right)\right\}.\tag{A.6}$$

Next, define $\varphi(x)$, $0\leq x\leq e^{-1}$, to be the smallest root t of the equation

$$t=e^{xt}.\tag{A.7}$$

It is well known that $\varphi(x)$ is well defined in $[0,e^{-1}]$, in particular, $\varphi(x)=1+x+O(x^2)$. The following lower bound will be useful in the proof of Proposition 6 in Appendix I.

Fact 3: With notations as above, suppose further that $\Omega+\Phi\leq e^{-1}$. Then, with φ defined by (A.7), [19, Th. 9] states that

$$\mathbb{P}\{S=0\}\geq\exp\{-\Delta\cdot\varphi(\Omega+\Phi)\}.\tag{A.8}$$

Appendix B

Proof of Theorem 3

Let us abbreviate $\mathcal{I}(m,m')\triangleq\mathcal{I}\left\{(\boldsymbol{x}_m,\boldsymbol{x}_{m'})\in\mathcal{T}(Q_{XX'})\right\}$, such that the enumerator $N(Q_{XX'})$ can also be written by

$$N(Q_{XX'})=\sum_{(m,m')\in[M]_*^2}\mathcal{I}(m,m'),\tag{B.1}$$

where the set $[M]_*^2$ is an abbreviation for the set $\{(m,m'):\ m,m'\in\{0,1,\ldots,M-1\},\ m\neq m'\}$.

Before proving Theorem 3, we start with the following series of partial results, that are going to be instrumental in proving Theorem 3.

Lemma 1: For any two pairs $(i,j),(i,k)\in[M]_*^2$, $j\neq k$,

$$\mathbb{E}[\mathcal{I}(i,j)\mathcal{I}(i,k)]\doteq\exp\{-2nI_Q(X;X')\}.\tag{B.2}$$

Proof: Since all codewords are independent, it follows by the method of types that

$$\begin{aligned}
&\mathbb{E}[\mathcal{I}(i,j)\mathcal{I}(i,k)] \\
&= \mathbb{P}\left\{(\boldsymbol{X}_i,\boldsymbol{X}_j)\in\mathcal{T}(Q_{XX'}),(\boldsymbol{X}_i,\boldsymbol{X}_k)\in\mathcal{T}(Q_{XX'})\right\} && \text{(B.3)}\\
&= \sum_{\boldsymbol{x}\in\mathcal{T}(Q_X)}\mathbb{P}\{\boldsymbol{X}_i=\boldsymbol{x}\} \\
&\quad\cdot\mathbb{P}\left\{(\boldsymbol{x},\boldsymbol{X}_j)\in\mathcal{T}(Q_{XX'}),(\boldsymbol{x},\boldsymbol{X}_k)\in\mathcal{T}(Q_{XX'})\right\} && \text{(B.4)}\\
&= \sum_{\boldsymbol{x}\in\mathcal{T}(Q_X)}\mathbb{P}\{\boldsymbol{X}_i=\boldsymbol{x}\}\cdot\mathbb{P}\left\{(\boldsymbol{x},\boldsymbol{X}_j)\in\mathcal{T}(Q_{XX'})\right\} \\
&\quad\cdot\mathbb{P}\left\{(\boldsymbol{x},\boldsymbol{X}_k)\in\mathcal{T}(Q_{XX'})\right\} && \text{(B.5)}\\
&\doteq \sum_{\boldsymbol{x}\in\mathcal{T}(Q_X)}\mathbb{P}\{\boldsymbol{X}_i=\boldsymbol{x}\}\cdot\exp\{-nI_Q(X;X')\} \\
&\quad\cdot\exp\{-nI_Q(X;X')\} && \text{(B.6)}\\
&= \exp\{-2nI_Q(X;X')\}, && \text{(B.7)}
\end{aligned}$$

where (B.5) is because $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$ are statistically independent. Lemma 1 is proved.

Now, we have the following Lemma, which proposes an upper bound on the probability of the lower tail in the case of TP type classes.

Lemma 2: Let $\epsilon > 0$ be given. Then, for any $Q_{XX'}$ such that $I_Q(X;X') \le 2R-\epsilon$,

$$\begin{aligned}
&\mathbb{P}\left\{N(Q_{XX'})\le e^{-n\epsilon}\cdot\mathbb{E}[N(Q_{XX'})]\right\} \\
&\quad\overset{\circ}{\le}\exp\left\{-\min\left(e^{n(2R-I_Q(X;X'))},e^{nR}\right)\right\}. && \text{(B.8)}
\end{aligned}$$

Proof: We use the result of Fact 1, that appears in Appendix A. In our case, we have $a = e^{-n\epsilon}$ and $\Delta \doteq e^{n(2R-I_Q(X;X'))}$, and it only remains to assess the quantities Θ and Ω. One can easily check that the indicator random variables $\mathcal{I}(i,j)$ and $\mathcal{I}(k,l)$ are independent as long as $i\ne k$ and $j\ne l$. Thus, we define our dependency graph in a way that each vertex (i,j) is connected to exactly $e^{nR}+e^{nR}-2$ vertices of the form (i,l), $l\ne j$ or (k,j), $k\ne i$. If the vertices (i,j) and (k,l) are connected, we denote it by $(i,j)\sim(k,l)$. Using the result of Lemma 1, we get that

$$\begin{aligned}
\Theta &= \frac{1}{2}\sum_{(i,j)\in[M]_*^2}\;\sum_{(k,l)\in[M]_*^2,(k,l)\sim(i,j)}\mathbb{E}[\mathcal{I}(i,j)\mathcal{I}(k,l)] && \text{(B.9)}\\
&\doteq \frac{1}{2}e^{2nR}\cdot(e^{nR}+e^{nR}-2)\cdot e^{-2nI_Q(X;X')} && \text{(B.10)}\\
&\doteq e^{n(3R-2I_Q(X;X'))}, && \text{(B.11)}
\end{aligned}$$

and

$$\begin{aligned}
\Omega &= \max_{(i,j)\in[M]_*^2}\sum_{(k,l)\in[M]_*^2,(k,l)\sim(i,j)}\mathbb{E}[\mathcal{I}(k,l)] && \text{(B.12)}\\
&\doteq (e^{nR}+e^{nR}-2)\cdot e^{-nI_Q(X;X')} && \text{(B.13)}\\
&\doteq e^{n(R-I_Q(X;X'))}. && \text{(B.14)}
\end{aligned}$$

Then,

$$\frac{\Delta}{6\Omega}\doteq\frac{e^{n(2R-I_Q(X;X'))}}{e^{n(R-I_Q(X;X'))}}=e^{nR}, \qquad \text{(B.15)}$$

and,

$$\begin{aligned}
\frac{\Delta^2}{8\Theta+2\Delta} &\doteq \frac{e^{n(4R-2I_Q(X;X'))}}{e^{n(3R-2I_Q(X;X'))}+e^{n(2R-I_Q(X;X'))}} && \text{(B.16)}\\
&= \frac{e^{n(2R-I_Q(X;X'))}}{e^{n(R-I_Q(X;X'))}+1} && \text{(B.17)}\\
&\doteq \frac{e^{n(2R-I_Q(X;X'))}}{e^{n[R-I_Q(X;X')]_+}}. && \text{(B.18)}
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{P}\left\{N(Q_{XX'})\le e^{-n\epsilon}\cdot\mathbb{E}[N(Q_{XX'})]\right\} \\
&\overset{\circ}{\le}\exp\left\{-\min\left(\frac{e^{n(2R-I_Q(X;X'))}}{e^{n[R-I_Q(X;X')]_+}},e^{nR}\right)\right\} && \text{(B.19)}\\
&=\exp\left\{-\min\left(e^{n(2R-I_Q(X;X'))},e^{nR}\right)\right\}. && \text{(B.20)}
\end{aligned}$$

Now, if $I_Q(X;X')\le R$, we get

$$\mathbb{P}\left\{N(Q_{XX'})\le e^{-n\epsilon}\cdot\mathbb{E}[N(Q_{XX'})]\right\}\overset{\circ}{\le}\exp\left\{-e^{nR}\right\}, \qquad \text{(B.21)}$$

and otherwise, if $R < I_Q(X;X')\le 2R-\epsilon$,

$$\begin{aligned}
&\mathbb{P}\left\{N(Q_{XX'})\le e^{-n\epsilon}\cdot\mathbb{E}[N(Q_{XX'})]\right\} \\
&\overset{\circ}{\le}\exp\left\{-e^{n(2R-I_Q(X;X'))}\right\} && \text{(B.22)}\\
&\le\exp\left\{-e^{n\epsilon}\right\}, && \text{(B.23)}
\end{aligned}$$

which completes the proof of Lemma 2.

Before moving on to the upper tail, we need the following lemma, proved in Appendix C.

Lemma 3: For any $k\in\mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}\left[N(Q_{XX'})^k\right] \\
&\dot{\le}\begin{cases}\exp\{nk\,(2R-I_Q(X;X'))\} & I_Q(X;X')<2R\\ \exp\{n\,(2R-I_Q(X;X'))\} & I_Q(X;X')>2R\end{cases}. && \text{(B.24)}
\end{aligned}$$

Concerning the upper tail, we have the following result.

Lemma 4: Let $\epsilon > 0$ be given. Then, for any $Q_{XX'}$ such that $I_Q(X;X')\le 2R$,

$$\mathbb{P}\left\{N(Q_{XX'})\ge e^{n\epsilon}\cdot\mathbb{E}[N(Q_{XX'})]\right\}\dot{\le}e^{-n\infty}. \qquad \text{(B.25)}$$

Proof: For any $k\in\mathbb{N}$, Markov's inequality and Lemma 3 implies that

$$\begin{aligned}
&\mathbb{P}\left\{N(Q_{XX'})\ge e^{n\epsilon}\cdot\mathbb{E}[N(Q_{XX'})]\right\} \\
&\le\inf_{k\in\mathbb{N}}\frac{\mathbb{E}[N(Q_{XX'})^k]}{e^{nk\epsilon}\cdot(\mathbb{E}[N(Q_{XX'})])^k} && \text{(B.26)}\\
&\dot{\le}\inf_{k\in\mathbb{N}}\frac{\exp\{nk\,(2R-I_Q(X;X'))\}}{e^{nk\epsilon}\cdot(\exp\{n\,(2R-I_Q(X;X'))\})^k} && \text{(B.27)}\\
&=\inf_{k\in\mathbb{N}}\exp\{-nk\epsilon\}, && \text{(B.28)}
\end{aligned}$$

thus,

$$\begin{aligned}
&\liminf_{n\to\infty}-\frac{1}{n}\log\mathbb{P}\left\{N(Q_{XX'})\ge e^{n\epsilon}\cdot\mathbb{E}[N(Q_{XX'})]\right\} \\
&\ge\sup_{k\in\mathbb{N}}k\epsilon=\infty, && \text{(B.29)}
\end{aligned}$$

which proves Lemma 4.

Next, we treat the TE type classes.

Lemma 5: Let $\epsilon > 0$ be given. Then, for any $Q_{XX'}$ such that $I_Q(X;X') \geq 2R$,

$$\mathbb{P}\left\{N(Q_{XX'}) \geq e^{n\epsilon}\right\} \stackrel{\cdot}{\leq} e^{-n\infty}. \quad \text{(B.30)}$$

Proof: For any $k \in \mathbb{N}$, Markov's inequality and Lemma 3 implies that

$$\begin{aligned}
&\mathbb{P}\left\{N(Q_{XX'}) \geq e^{n\epsilon}\right\} \\
&\leq \inf_{k\in\mathbb{N}} \frac{\mathbb{E}[N(Q_{XX'})^k]}{e^{nk\epsilon}} && \text{(B.31)}\\
&\stackrel{\cdot}{\leq} \inf_{k\in\mathbb{N}} \frac{\exp\left\{n\left(2R - I_Q(X;X')\right)\right\}}{e^{nk\epsilon}} && \text{(B.32)}\\
&= \inf_{k\in\mathbb{N}} \exp\left\{-n\left(I_Q(X;X') - 2R + k\epsilon\right)\right\}, && \text{(B.33)}
\end{aligned}$$

and hence,

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{P}\left\{N(Q_{XX'}) \geq e^{n\epsilon}\right\} \\
&\geq \sup_{k\in\mathbb{N}} \left\{I_Q(X;X') - 2R + k\epsilon\right\} = \infty, && \text{(B.34)}
\end{aligned}$$

which completes the proof of Lemma 5. Furthermore, we have

Lemma 6: For any $Q_{XX'}$ such that $I_Q(X;X') \geq 2R$,

$$\mathbb{P}\left\{N(Q_{XX'}) \geq 1\right\} \doteq \exp\{n(2R - I_Q(X;X'))\}. \quad \text{(B.35)}$$

Proof: An upper bound simply follows from Markov's inequality:

$$\begin{aligned}
\mathbb{P}\left\{N(Q_{XX'}) \geq 1\right\} &\leq \mathbb{E}[N(Q_{XX'})] && \text{(B.36)}\\
&\doteq \exp\{n(2R - I_Q(X;X'))\}. && \text{(B.37)}
\end{aligned}$$

For the lower bound, we use Fact 2 from Appendix A. Similarly to (B.15) and (B.16), we have

$$\frac{\Delta^2}{8\Theta} \doteq \frac{e^{n(4R-2I_Q(X;X'))}}{e^{n(3R-2I_Q(X;X'))}} = e^{nR}, \quad \text{(B.38)}$$

and,

$$\frac{\Delta}{6\Omega} \doteq \frac{e^{n(2R-I_Q(X;X'))}}{e^{n(R-I_Q(X;X'))}} = e^{nR}. \quad \text{(B.39)}$$

Now, since $I_Q(X;X') \geq 2R$,

$$\begin{aligned}
&\mathbb{P}\left\{N(Q_{XX'}) = 0\right\} \\
&\leq \exp\left\{-\min\left(e^{nR}, e^{nR}, \frac{1}{2}\cdot e^{n(2R-I_Q(X;X'))}\right)\right\} && \text{(B.40)}\\
&= \exp\left\{-\frac{1}{2}\cdot e^{n(2R-I_Q(X;X'))}\right\} && \text{(B.41)}\\
&\leq 1 - \frac{1}{2}\cdot e^{n(2R-I_Q(X;X'))} + \frac{1}{8}\cdot e^{n(4R-2I_Q(X;X'))}, && \text{(B.42)}
\end{aligned}$$

where (B.42) is due to the fact that for $t \geq 0$, $e^{-t} \leq 1 - t + t^2/2$, and so,

$$\begin{aligned}
\mathbb{P}\left\{N(Q_{XX'}) \geq 1\right\} &= 1 - \mathbb{P}\left\{N(Q_{XX'}) = 0\right\} && \text{(B.43)}\\
&\geq \frac{1}{2}\cdot \exp\{n(2R - I_Q(X;X'))\} \\
&\quad - \frac{1}{8}\cdot \exp\{n(4R - 2I_Q(X;X'))\} && \text{(B.44)}\\
&\doteq \exp\{n(2R - I_Q(X;X'))\}, && \text{(B.45)}
\end{aligned}$$

which is compatible with the above upper bound, proving Lemma 6.

Proof of Theorem 3: Let us abbreviate $I_Q = I_Q(X;X')$. We use the results of Lemmas 2, 4, 5, and 6, and get the following exponential rate of decay for $\mathbb{P}\left\{N(Q_{XX'}) \geq e^{ns}\right\}$:

$$\begin{aligned}
&E(R,Q,s) \\
&= \begin{cases} I_Q - 2R & I_Q \geq 2R, s \leq 0 \\ \infty & I_Q \geq 2R, s > 0 \\ 0 & I_Q \leq 2R, s \leq 2R - I_Q \\ \infty & I_Q \leq 2R, s > 2R - I_Q \end{cases} && \text{(B.46)}\\
&= \begin{cases} [I_Q - 2R]_+ & I_Q \geq 2R, s \leq [2R - I_Q]_+ \\ \infty & I_Q \geq 2R, s > [2R - I_Q]_+ \\ [I_Q - 2R]_+ & I_Q \leq 2R, s \leq [2R - I_Q]_+ \\ \infty & I_Q \leq 2R, s > [2R - I_Q]_+ \end{cases} && \text{(B.47)}\\
&= \begin{cases} [I_Q - 2R]_+ & [2R - I_Q]_+ \geq s \\ \infty & [2R - I_Q]_+ < s \end{cases}, && \text{(B.48)}
\end{aligned}$$

which proves Theorem 3.

APPENDIX C

Proof of Lemma 3

For a set of indices $\mathcal{J}$ let us denote $\mathcal{J}_*^2 = \{(j,j') \in \mathcal{J}^2 : j \neq j'\}$. Recall that $\mathcal{I}(m,m') = \mathcal{I}\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'})\}$ and $N(Q_{XX'}) = \sum_{(m,m')\in[M]_*^2} \mathcal{I}(m,m')$. We show by induction that

$$\mathbb{E}\left[N(Q_{XX'})^k\right] \stackrel{\cdot}{\leq} \begin{cases} e^{nk(2R-I)} & I < 2R \\ e^{-n(I-2R)} & I > 2R \end{cases}, \quad \text{(C.1)}$$

where I is a shorthand notation for $I_Q(X;X')$. This clearly holds for $k = 1$ by linearity of expectation. We assume it holds up to $k-1$ and show this for k.

<u>Proof for k</u>: Assume that $\{(m_i, m_i')\}_{i=1}^{k-1}$ are given, where $(m_i, m_i') \in [M]_*^2$ for all $i \in [k-1]$. Let $\mathcal{M}_{k-1} = \bigcup_{i=1}^{k-1}\{\{m_i\} \cup \{m_i'\}\}$ be the set of indices of the $k-1$ pairs of codeword indices $\{(m_i, m_i')\}_{i=1}^{k-1}$. We condition on all these codewords, and then compute expectation w.r.t. all other codewords. For any fixed k, the number of codewords in the first $k-1$ indicators is negligible to the number of all other codewords. Specifically, $|\mathcal{M}_{k-1}| \leq 2(k-1)$ holds. Now,

$$\begin{aligned}
&\sum_{(m_k,m_k')\in[M]_*^2} \mathcal{I}(m_k, m_k') \\
&= \sum_{(m_k,m_k')\in([M]\setminus\mathcal{M}_{k-1})_*^2} \mathcal{I}(m_k, m_k') \\
&\quad + \sum_{m_k\in\mathcal{M}_{k-1}} \sum_{m_k'\in[M]\setminus\mathcal{M}_{k-1}} \left(\mathcal{I}(m_k, m_k') + \mathcal{I}(m_k', m_k)\right) \\
&\quad + \sum_{(m_k,m_k')\in(\mathcal{M}_{k-1})_*^2} \mathcal{I}(m_k, m_k'). && \text{(C.2)}
\end{aligned}$$

By (C.2), linearity of expectation, the independence of codewords assumption, and the trivial fact that $\mathcal{I}(m_k, m_k') \leq 1$,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{(m_k,m_k')\in[M]_*^2} \mathcal{I}(m_k, m_k') \,\middle|\, \{\boldsymbol{X}_l\}_{l\in\mathcal{M}_{k-1}}\right] \\
&\stackrel{\cdot}{\leq} e^{n(2R-I)} + 4(k-1)e^{n(R-I)} + 4(k-1)^2 && \text{(C.3)}\\
&\doteq \max\{e^{n(2R-I)}, 1\}. && \text{(C.4)}
\end{aligned}$$

Now,

$$\begin{aligned}
&\mathbb{E}\left[N(Q_{XX'})^k\right] \\
&= \sum_{\left\{\substack{(m_i,m_i')\in[M]_*^2,\\ 1\le i\le k}\right\}} \mathbb{E}\left[\prod_{i=1}^{k}\mathcal{I}(m_i,m_i')\right] && \text{(C.5)}\\
&= \sum_{\left\{\substack{(m_i,m_i')\in[M]_*^2,\\ 1\le i\le k-1}\right\}} \mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i') \cdot\left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\right]. && \text{(C.6)}
\end{aligned}$$

The expectation in (C.6) is given by

$$\begin{aligned}
&\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\Big|\{\boldsymbol{X}_l\}_{l\in\mathcal{M}_{k-1}}\right]\right] && \text{(C.7)}\\
&=\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\mathbb{E}\left[\left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\Big|\{\boldsymbol{X}_l\}_{l\in\mathcal{M}_{k-1}}\right]\right] && \text{(C.8)}\\
&\stackrel{\cdot}{\le}\max\{e^{n(2R-I)},1\}\cdot\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\right], && \text{(C.9)}
\end{aligned}$$

where (C.8) is due to the fact that upon conditioning on $\{\boldsymbol{X}_l\}_{l\in\mathcal{M}_{k-1}}$, $\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')$ is fixed, and (C.9) follows from (C.4). Substituting it back into (C.6) and using the induction assumption provides

$$\begin{aligned}
&\mathbb{E}\left[N(Q_{XX'})^k\right]\\
&\stackrel{\cdot}{\le}\max\{e^{n(2R-I)},1\}\sum_{\left\{\substack{(m_i,m_i')\in[M]_*^2,\\ 1\le i\le k-1}\right\}}\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\right] && \text{(C.10)}\\
&=\max\{e^{n(2R-I)},1\}\cdot\mathbb{E}\left[(N(Q_{XX'}))^{k-1}\right] && \text{(C.11)}\\
&\stackrel{\cdot}{\le}\max\{e^{n(2R-I)},1\}\cdot\begin{cases}e^{n(k-1)(2R-I)} & I<2R\\ e^{-n(I-2R)} & I>2R\end{cases} && \text{(C.12)}\\
&=\begin{cases}e^{nk(2R-I)} & I<2R\\ e^{-n(I-2R)} & I>2R\end{cases}. && \text{(C.13)}
\end{aligned}$$

Thus, Lemma 3 is proved.

Appendix D

Proof of Proposition 1

The monotonicity is straightforward, and follows the fact that $\mathcal{L}(R,E_0)$ and $\mathcal{M}(R,E_0)$, defined in (26) and (27), respectively, become larger when E_0 grows. In order to show the fourth item, observe that when $E_0 < E_0^{\text{min}}$, the set $\mathcal{L}(R,E_0)$ is empty. As for the second item, we seek a condition on E_0 such that $E_{\text{lt}}^{\text{ub}}(R,E_0)>0$:

$$\min_{Q_{XX'}\in\mathcal{L}(R,E_0)}[I_Q(X;X')-2R]_+>0. \tag{D.1}$$

Explicitly,

$$\min_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ [2R-I_Q(X;X')]_+\ge\Gamma(Q_{XX'},R)+R-E_0\}}[I_Q(X;X')-2R]_+>0, \tag{D.2}$$

and by using the identity $\min_{\{Q:\ g(Q)\le 0\}}f(Q)=\min_Q\sup_{s\ge0}\{f(Q)+s\cdot g(Q)\}$, it can also be written as

$$\begin{aligned}
\min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\sup_{s\ge0}\{&s\cdot(\Gamma(Q_{XX'},R)+R-E_0\\
&-[2R-I_Q(X;X')]_+)+[I_Q(X;X')-2R]_+\}>0, && \text{(D.3)}
\end{aligned}$$

which means that for every $Q_{XX'}\in\mathcal{Q}(Q_X)$ there exists some $s\ge0$, such that

$$\begin{aligned}
&s\cdot(\Gamma(Q_{XX'},R)+R-E_0-[2R-I_Q(X;X')]_+)\\
&\qquad+[I_Q(X;X')-2R]_+>0, && \text{(D.4)}
\end{aligned}$$

or equivalently,

$$\begin{aligned}
E_0<{}&\Gamma(Q_{XX'},R)+R-[2R-I_Q(X;X')]_+\\
&+\frac{[I_Q(X;X')-2R]_+}{s}. && \text{(D.5)}
\end{aligned}$$

Thus,

$$\begin{aligned}
E_0&<\min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\sup_{s\ge0}\Big\{\Gamma(Q_{XX'},R)+R\\
&\qquad-[2R-I_Q(X;X')]_++\frac{[I_Q(X;X')-2R]_+}{s}\Big\} && \text{(D.6)}\\
&=\min_{Q_{XX'}\in\mathcal{Q}(Q_X)}\Big[\Gamma(Q_{XX'},R)+R-[2R-I_Q(X;X')]_+\\
&\qquad+\begin{cases}0 & I_Q(X;X')\le 2R\\ \infty & I_Q(X;X')>2R\end{cases}\Big] && \text{(D.7)}\\
&=\min_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\le2R\}}\{\Gamma(Q_{XX'},R)+R\\
&\qquad-[2R-I_Q(X;X')]_+\} && \text{(D.8)}\\
&=\min_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\le2R\}}\{\Gamma(Q_{XX'},R)\\
&\qquad+I_Q(X;X')-R\} && \text{(D.9)}\\
&=E_{\text{trc}}(R), && \text{(D.10)}
\end{aligned}$$

where the ∞ in (D.7) is because the maximizing $s\ge0$ in (D.6) when $I_Q(X;X')>2R$ is $s^*=0$. The proof of the third item is very similar to the proof of the second item and hence omitted.

Appendix E

Proof of Proposition 3

The monotonicity is immediate, since both $\mathcal{V}(R,E_0)$ and $\mathcal{U}(R,E_0)$, defined in (42) and (43), respectively, become larger when E_0 grows. In order to show the second item, we seek a condition on E_0 such that $E_{\text{ut}}^{\text{lb}}(R,E_0)>0$:

$$\max_{Q_{XX'}\in\mathcal{U}(R,E_0)}\{2R-I_Q(X;X')\}>0. \tag{E.1}$$

Explicitly,

$$\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\leq 2R,\ \Gamma(Q_{XX'},R)+I_Q(X;X')-R\leq E_0\}}\{2R-I_Q(X;X')\}>0, \quad \text{(E.2)}$$

and thanks to the fact that $\max_{\{Q:\ g(Q)\geq 0\}} f(Q) = \max_Q \inf_{\mu\geq 0}\{f(Q)+\mu\cdot g(Q)\}$, it can also be written as

$$\max_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\leq 2R\}}\inf_{\mu\geq 0}\{2R-I_Q(X;X') + \mu\cdot(E_0-\Gamma(Q_{XX'},R)-I_Q(X;X')+R)\}>0, \quad \text{(E.3)}$$

or, equivalently,

$$\exists Q_{XX'}\in\mathcal{Q}(Q_X) \text{ s.t. } I_Q(X;X')\leq 2R,\quad \forall\mu\geq 0: \mu\cdot E_0 > I_Q(X;X')-2R + \mu\cdot(\Gamma(Q_{XX'},R)+I_Q(X;X')-R), \quad \text{(E.4)}$$

or,

$$\begin{aligned} E_0 &> \min_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\leq 2R\}}\sup_{\mu\geq 0}\left\{\frac{I_Q(X;X')-2R}{\mu} + \Gamma(Q_{XX'},R)+I_Q(X;X')-R\right\} && \text{(E.5)}\\ &= \min_{\{Q_{XX'}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\leq 2R\}}\{\Gamma(Q_{XX'},R) + I_Q(X;X')-R\} && \text{(E.6)}\\ &= E_{\text{trc}}(R), && \text{(E.7)}\end{aligned}$$

where (E.6) is because the maximizing $\mu\geq 0$ in (E.5) is $\mu^*=\infty$, since $I_Q(X;X')\leq 2R$. The proof of the third item is very similar to the proof of the second item and hence omitted.

Appendix F

Proof of Corollary 1

The probability of any codebook in the ensemble is given asymptotically by $\exp\{-nH_Q(X)e^{nR}\}$, hence, in order to assure that a code exists, we demand that

$$\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n)\geq E_0\right\} > \exp\{-nH_Q(X)e^{nR}\}. \quad \text{(F.1)}$$

Now, the lower bound of Theorem 2 reads

$$\mathbb{P}\left\{-\frac{1}{n}\log P_{\text{e}}(\mathcal{C}_n)\geq E_0\right\} \overset{\circ}{\geq} \exp\left\{-\exp\left\{n\cdot\max_{Q_{XX'}\in\mathcal{U}(R,E_0)}\{2R-I_Q(X;X')\}\right\}\right\}, \quad \text{(F.2)}$$

thus (F.1) will obviously be satisfied if

$$\max_{Q_{XX'}\in\mathcal{U}(R,E_0)}\{2R-I_Q(X;X')\}<R, \quad \text{(F.3)}$$

or, equivalently,

$$\min_{Q_{XX'}\in\mathcal{U}(R,E_0)} I_Q(X;X')>R, \quad \text{(F.4)}$$

which is exactly (I.19). Then, following some algebraic work, that can be found in (I.20)–(I.30), we found that (F.4) is equivalent to $E_0<E_{\text{ex}}(R)$.

Appendix G

Proof of Proposition 4

For a set of indices $\mathcal{J}$ let us denote $\mathcal{J}_*^2=\{(j,j')\in\mathcal{J}^2:\ j\neq j'\}$. Recall that $\mathcal{I}(m,m')=\mathcal{I}\{(\boldsymbol{X}_m,\boldsymbol{X}_{m'})\in\mathcal{T}(Q_{XX'})\}$ and $N(Q_{XX'})=\sum_{(m,m')\in[M]_*^2}\mathcal{I}(m,m')$. Let us abbreviate $\mathcal{I}(m)=\mathcal{I}\{(\boldsymbol{X}_m,\boldsymbol{y})\in\mathcal{T}(Q_{XY})\}$, such that

$$N_{\boldsymbol{y}}(Q_{XY})=\sum_{m\in[M]}\mathcal{I}(m). \quad \text{(G.1)}$$

Recall the definition of $F(S,Q_{UV},j)$ in (115). We show by induction that

$$\mathbb{E}\left[N_{\boldsymbol{y}}(Q_{XY})^l N(Q_{XX'})^k\right] \overset{\cdot}{\leq} F(R,Q_{XY},l)\cdot F(2R,Q_{XX'},k). \quad \text{(G.2)}$$

Checking for $k=l=1$: Note that due to the symmetry of the random draw over the type class:

$$\begin{aligned}\mathbb{E}[\mathcal{I}(m,m')\mathcal{I}(m)] &= \mathbb{E}[\mathcal{I}(m)\mathbb{E}[\mathcal{I}(m,m')\mid \boldsymbol{X}_m]] && \text{(G.3)}\\ &= \mathbb{E}[\mathcal{I}(m)]\cdot\mathbb{E}[\mathcal{I}(m,m')] && \text{(G.4)}\end{aligned}$$

and similarly, $\mathbb{E}[\mathcal{I}(m,m')\mathcal{I}(m')]=\mathbb{E}[\mathcal{I}(m')]\cdot\mathbb{E}[\mathcal{I}(m,m')]$. Thus, for $k=l=1$:

$$\begin{aligned}&\mathbb{E}[N_{\boldsymbol{y}}(Q_{XY})N(Q_{XX'})]\\ &= \sum_{(m,m')\in[M]_*^2}\sum_{r\in[M]}\mathbb{E}[\mathcal{I}(m,m')\mathcal{I}(r)] && \text{(G.5)}\\ &= \sum_{(m,m')\in[M]_*^2}\left(\sum_{r\in[M]\setminus\{m,m'\}}\mathbb{E}[\mathcal{I}(m,m')]\,\mathbb{E}[\mathcal{I}(r)] + \mathbb{E}[\mathcal{I}(m,m')\mathcal{I}(m)]+\mathbb{E}[\mathcal{I}(m,m')\mathcal{I}(m')]\right) && \text{(G.6)}\\ &= \sum_{(m,m')\in[M]_*^2}\sum_{r\in[M]}\mathbb{E}[\mathcal{I}(m,m')]\,\mathbb{E}[\mathcal{I}(r)] && \text{(G.7)}\\ &\doteq e^{n(2R-I_Q(X;X'))}\cdot e^{n(R-I_Q(X;Y))}. && \text{(G.8)}\end{aligned}$$

Induction assumption: Assume that (G.2) holds up for some $(k-1,l-1)$. We show by two inductive steps that this holds for $(k,l-1)$ and $(k-1,l)$ and thus for any (k,l).

Proof for $(k,l-1)$: Assume that $\{(m_i,m_i')\}_{i=1}^{k-1}$ and $\{r_j\}_{j=1}^{l-1}$ are given, where $(m_i,m_i')\in[M]_*^2$ for all $i\in[k-1]$, and $r_j\in[M]$ for all $j\in[l-1]$. Let $\mathcal{M}_{k-1,l-1}=\bigcup_{i=1}^{k-1}\{\{m_i\}\cup\{m_i'\}\}\cup\bigcup_{j=1}^{l-1}\{r_j\}$ be the set of indices of the $k-1$ pairs of codeword indices $\{(m_i,m_i')\}_{i=1}^{k-1}$ and of the $l-1$ codeword indices $\{r_j\}_{j=1}^{l-1}$. Clearly $|\mathcal{M}_{k-1,l-1}|\leq 2(k-1)+l-1\triangleq c_{k-1,l-1}$ holds. Now,

$$\begin{aligned}&\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\\ &= \sum_{(m_k,m_k')\in([M]\setminus\mathcal{M}_{k-1,l-1})_*^2}\mathcal{I}(m_k,m_k')\\ &\quad+\sum_{m_k\in\mathcal{M}_{k-1,l-1}}\sum_{m_k'\in[M]\setminus\mathcal{M}_{k-1,l-1}}(\mathcal{I}(m_k,m_k')+\mathcal{I}(m_k',m_k))\\ &\quad+\sum_{(m_k,m_k')\in(\mathcal{M}_{k-1,l-1})_*^2}\mathcal{I}(m_k,m_k'). && \text{(G.9)}\end{aligned}$$

By (G.9), linearity of expectation, the independence of codewords assumption, and the fact that $\mathcal{I}(m_k, m_k') \le 1$,

$$\mathbb{E}\left[\sum_{(m_k,m_k')\in[M]_*^2} \mathcal{I}(m_k, m_k')\middle|\{\boldsymbol{X}_s\}_{s\in\mathcal{M}_{k-1,l-1}}\right] \dot{\le} e^{n(2R-I_Q(X;X'))} + 2c_{k-1,l-1}e^{n(R-I_Q(X;X'))} + c_{k-1,l-1}^2 \tag{G.10}$$

$$\doteq \max\{e^{n(2R-I_Q(X;X'))}, 1\}. \tag{G.11}$$

Next,

$$\mathbb{E}\left[N_{\boldsymbol{y}}(Q_{XY})^{l-1}N(Q_{XX'})^k\right] = \sum_{\left\{\substack{(m_i,m_i')\in[M]_*^2,\\ 1\le i\le k}\right\}} \sum_{\left\{\substack{r_j\in[M],\\ 1\le j\le l-1}\right\}} \mathbb{E}\left[\prod_{i=1}^{k}\mathcal{I}(m_i,m_i')\prod_{j=1}^{l-1}\mathcal{I}(r_j)\right] \tag{G.12}$$

$$= \sum_{\left\{\substack{(m_i,m_i')\in[M]_*^2,\\ 1\le i\le k-1}\right\}} \sum_{\left\{\substack{r_j\in[M],\\ 1\le j\le l-1}\right\}} \mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\prod_{j=1}^{l-1}\mathcal{I}(r_j) \left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\right]. \tag{G.13}$$

The expectation in (G.13) is given by

$$\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\prod_{j=1}^{l-1}\mathcal{I}(r_j)\left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\prod_{j=1}^{l-1}\mathcal{I}(r_j) \left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\middle|\{\boldsymbol{X}_s\}_{s\in\mathcal{M}_{k-1,l-1}}\right]\right] \tag{G.14}$$

$$= \mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\prod_{j=1}^{l-1}\mathcal{I}(r_j) \cdot\mathbb{E}\left[\left(\sum_{(m_k,m_k')\in[M]_*^2}\mathcal{I}(m_k,m_k')\right)\middle|\{\boldsymbol{X}_s\}_{s\in\mathcal{M}_{k-1,l-1}}\right]\right] \tag{G.15}$$

$$\dot{\le} \max\{e^{n(2R-I_Q(X;X'))}, 1\} \cdot\mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\prod_{j=1}^{l-1}\mathcal{I}(r_j)\right], \tag{G.16}$$

where (G.15) is thanks to the conditioning on $\{\boldsymbol{X}_s\}_{s\in\mathcal{M}_{k-1,l-1}}$, and (G.16) is due to (G.11). Substituting it back into (G.13) and using the induction assumption provides

$$\mathbb{E}\left[N_{\boldsymbol{y}}(Q_{XY})^{l-1}N(Q_{XX'})^k\right] \dot{\le} \max\{e^{n(2R-I_Q(X;X'))}, 1\} \sum_{\left\{\substack{(m_i,m_i')\in[M]_*^2,\\ 1\le i\le k-1}\right\}} \sum_{\left\{\substack{r_j\in[M],\\ 1\le j\le l-1}\right\}} \mathbb{E}\left[\prod_{i=1}^{k-1}\mathcal{I}(m_i,m_i')\cdot\prod_{j=1}^{l-1}\mathcal{I}(r_j)\right] \tag{G.17}$$

$$= \max\{e^{n(2R-I_Q(X;X'))}, 1\} \cdot\mathbb{E}\left[N_{\boldsymbol{y}}(Q_{XY})^{l-1}N(Q_{XX'})^{k-1}\right] \tag{G.18}$$

$$\dot{\le} \max\{e^{n(2R-I_Q(X;X'))}, 1\} \cdot F(R, Q_{XY}, l-1)\cdot F(2R, Q_{XX'}, k-1) \tag{G.19}$$

$$= F(R, Q_{XY}, l-1)\cdot F(2R, Q_{XX'}, k), \tag{G.20}$$

which completes the proof of the first inductive step. The proof of the second inductive step follows exactly the same lines and hence omitted. The proof of Proposition 4 is complete.

Appendix H

Proof of Proposition 5

By the union bound,

$$\mathbb{P}\left\{\hat{\mathcal{B}}_n(\sigma)\right\} = \mathbb{P}\left\{\bigcup_{m=0}^{M-1}\bigcup_{m'\ne m}\bigcup_{\boldsymbol{y}\in\mathcal{Y}^n}\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y})\right\} \tag{H.1}$$

$$\le \sum_{m=0}^{M-1}\sum_{m'\ne m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\left\{\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y})\right\}. \tag{H.2}$$

Now,

$$\mathbb{P}\left\{\hat{\mathcal{B}}_n(\sigma,m,m',\boldsymbol{y})\right\} = \mathbb{P}\left\{\sum_{\tilde{m}\in\{0,1,\ldots,M-1\}\setminus\{m,m'\}} e^{ng(\hat{P}_{\boldsymbol{X}_{\tilde{m}}\boldsymbol{y}})} \ge e^{n\cdot(\beta(R,Q_Y)+\sigma)}\right\} \tag{H.3}$$

$$= \mathbb{P}\left\{\sum_{Q_{XY}} N(Q_{XY})e^{ng(Q_{XY})} \ge e^{n\cdot(\beta(R,Q_Y)+\sigma)}\right\} \tag{H.4}$$

$$\doteq \sum_{Q_{XY}}\mathbb{P}\left\{N(Q_{XY}) \ge e^{n(\beta(R,Q_Y)+\sigma-g(Q_{XY}))}\right\} \tag{H.5}$$

$$= \sum_{\{Q_{XY}:\ I_Q(X;Y)\le R\}}\mathbb{P}\left\{N(Q_{XY}) \ge e^{n(\beta(R,Q_Y)+\sigma-g(Q_{XY}))}\right\} + \sum_{\{Q_{XY}:\ I_Q(X;Y)> R\}}\mathbb{P}\left\{N(Q_{XY}) \ge e^{n(\beta(R,Q_Y)+\sigma-g(Q_{XY}))}\right\}, \tag{H.6}$$

where (H.3) is due to the definition of $Z_{mm'}(\boldsymbol{y})$ in (74), in (H.4) we introduced the type class enumerator $N(Q_{XY})$, which is the number of codewords in $\mathcal{C}_n$, other than $\boldsymbol{x}_m$ and $\boldsymbol{x}_{m'}$, that have a joint composition Q_{XY} together with $\boldsymbol{y}$, and where (H.5) is due to the SME. The first summand of (H.6) is upper–bounded by

$$\mathbb{P}\left\{N(Q_{XY}) \ge \exp\left\{n\left(\beta(R,Q_Y)+\sigma-g(Q_{XY})\right)\right\}\right\} = \mathbb{P}\left\{N(Q_{XY}) \ge \exp\left\{n\left(\sigma+\beta(R,Q_Y)-g(Q_{XY}) - [R-I_Q(X;Y)]_+ + [R-I_Q(X;Y)]_+\right)\right\}\right\} \tag{H.7}$$

$$\le \mathbb{P}\left\{N(Q_{XY}) \ge \exp\left\{n\left(\sigma+[R-I_Q(X;Y)]_+\right)\right\}\right\} \tag{H.8}$$

$$= \mathbb{P}\left\{N(Q_{XY}) \ge e^{n(\sigma+R-I_Q(X;Y))}\right\} \tag{H.9}$$

$$\le \exp\left\{-e^{nR}D(e^{-n[R-(\sigma+R-I_Q(X;Y))]}\|e^{-nI_Q(X;Y)})\right\} \tag{H.10}$$

$$= \exp\left\{-e^{nR} D(e^{-n(I_Q(X;Y)-\sigma)} \| e^{-nI_Q(X;Y)})\right\} \quad \text{(H.11)}$$

$$< \exp\left\{-e^{nR} \cdot e^{-n(I_Q(X;Y)-\sigma)} \cdot \left(\ln \frac{e^{-n(I_Q(X;Y)-\sigma)}}{e^{-nI_Q(X;Y)}} - 1\right)\right\} \quad \text{(H.12)}$$

$$= \exp\left\{-e^{n(R-I_Q(X;Y)+\sigma)} \cdot (n\sigma - 1)\right\} \quad \text{(H.13)}$$

$$\leq \exp\left\{-e^{n\sigma}\right\}. \quad \text{(H.14)}$$

In (H.8), we use the definition of $\beta(R, Q_Y)$ in (22), which implies that $\beta(R, Q_Y) \geq g(Q_{XY}) + [R - I_Q(X;Y)]_+$, and for (H.9), recall that $R \geq I_Q(X;Y)$. Step (H.10) is according to Chernoff's bound [16, Appendix], [12, Appendix B], (H.12) is due to the following lower bound to the binary divergence [22, Sec. 6.3, p. 167]

$$D(a\|b) > a\left(\ln \frac{a}{b} - 1\right), \quad \text{(H.15)}$$

and (H.14) is true since $R \geq I_Q(X;Y)$. Similarly, for the second summand of (H.6), we have

$$\mathbb{P}\left\{N(Q_{XY}) \geq \exp\left\{n\left(\beta(R,Q_Y) + \sigma - g(Q_{XY})\right)\right\}\right\}$$
$$\leq \mathbb{P}\left\{N(Q_{XY}) \geq \exp\left\{n\left(\sigma + [R - I_Q(X;Y)]_+\right)\right\}\right\} \quad \text{(H.16)}$$

$$= \mathbb{P}\left\{N(Q_{XY}) \geq e^{n\sigma}\right\} \quad \text{(H.17)}$$

$$\leq \exp\left\{-e^{nR} D(e^{-n(R-\sigma)} \| e^{-nI_Q(X;Y)})\right\} \quad \text{(H.18)}$$

$$< \exp\left\{-e^{nR} \cdot e^{-n(R-\sigma)} \cdot \left(\ln \frac{e^{-n(R-\sigma)}}{e^{-nI_Q(X;Y)}} - 1\right)\right\} \quad \text{(H.19)}$$

$$= \exp\left\{-e^{n\sigma} \cdot [n(I_Q(X;Y) - R + \sigma) - 1]\right\} \quad \text{(H.20)}$$

$$\leq \exp\left\{-e^{n\sigma}\right\}, \quad \text{(H.21)}$$

where (H.16) is true for the same reason as (H.8), (H.17) is because $I_Q(X;Y) > R$, (H.18) is again due to Chernoff's bound, (H.19) is true thanks to (H.15), and (H.21) is due to $I_Q(X;Y) - R + \sigma > 0$. Hence, we conclude that for every $\sigma > 0$

$$\mathbb{P}\{\hat{\mathcal{B}}_n(\sigma, m, m', \boldsymbol{y})\}$$
$$= \mathbb{P}\{Z_{mm'}(\boldsymbol{y}) \geq \exp\{n \cdot (\beta(R, Q_Y) + \sigma)\}\} \quad \text{(H.22)}$$

$$\overset{\circ}{\leq} \exp\left\{-e^{n\sigma}\right\}, \quad \text{(H.23)}$$

and so, continuing from (H.2), this means that

$$\mathbb{P}\left\{\hat{\mathcal{B}}_n(\sigma)\right\} \overset{\circ}{\leq} \sum_{m=0}^{M-1} \sum_{m' \neq m} \sum_{\boldsymbol{y} \in \mathcal{Y}^n} \exp\left\{-e^{n\sigma}\right\} \quad \text{(H.24)}$$

$$\overset{\circ}{=} \exp\left\{-e^{n\sigma}\right\}, \quad \text{(H.25)}$$

which completes the proof of the proposition.

APPENDIX I

Proof of Proposition 6

First, note that

$$\mathcal{F}_0 = \left\{\sum_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} N(Q_{XX'}) = 0\right\}. \quad \text{(I.1)}$$

Let us define

$$N(\mathcal{A}_1 \cup \mathcal{A}_2) \triangleq \sum_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} N(Q_{XX'}), \quad \text{(I.2)}$$

and the binary random variables

$$\mathcal{I}(m, m', Q_{XX'}) \triangleq \mathcal{I}\left\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'})\right\}, \quad \text{(I.3)}$$

such that,

$$N(\mathcal{A}_1 \cup \mathcal{A}_2) = \sum_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \sum_{m=0}^{M-1} \sum_{m' \neq m} \mathcal{I}(m, m', Q_{XX'}). \quad \text{(I.4)}$$

In order to use Fact 3 that appears in Appendix A, let us first define an appropriate dependency graph. One can easily check that the indicator random variables $\mathcal{I}(i, j, Q)$ and $\mathcal{I}(k, l, \tilde{Q})$ are independent as long as $i \neq k$, $j \neq l$, and $Q \neq \tilde{Q}$. Thus, we define our dependency graph in a way that each vertex (i, j, Q) is connected to exactly $e^{nR} - 1$ vertices of the form (k, j, Q), $k \neq i$, to $e^{nR} - 1$ vertices of the form (i, l, Q), $l \neq j$, and to exactly $|\mathcal{A}_1 \cup \mathcal{A}_2| - 1$ vertices of the form $(i, j, \tilde{Q})$, $\tilde{Q} \neq Q$. Let us now examine the quantities Δ, Ω, and Φ. First,

$$\Delta = \mathbb{E}[N(\mathcal{A}_1 \cup \mathcal{A}_2)] \quad \text{(I.5)}$$

$$= \sum_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \mathbb{E}[N(Q_{XX'})] \quad \text{(I.6)}$$

$$\doteq \sum_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} e^{n \cdot (2R - I_Q(X;X'))} \quad \text{(I.7)}$$

$$\doteq \max_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} e^{n \cdot (2R - I_Q(X;X'))} \quad \text{(I.8)}$$

$$= \exp\left\{n \cdot \max_{Q_{XX'} \in \mathcal{A}_1 \cup \mathcal{A}_2} \{2R - I_Q(X;X')\}\right\} \quad \text{(I.9)}$$

$$= \exp\left\{n \cdot \max_{Q_{XX'} \in \mathcal{A}_2} \{2R - I_Q(X;X')\}\right\}, \quad \text{(I.10)}$$

where the last equality follows from the definitions of $\mathcal{A}_1$ and $\mathcal{A}_2$ and the assumption that $\mathcal{A}_2$ is nonempty. Regarding the quantity $\Omega_{i,j,Q}$ of (A.2), notice that it actually depends only on Q. Thus, for some $Q \in \mathcal{A}_1 \cup \mathcal{A}_2$,

$$\Omega_Q \doteq (e^{nR} + e^{nR} - 2) \cdot e^{-nI_Q(X;X')} + \sum_{\tilde{Q} \in \mathcal{A}_1 \cup \mathcal{A}_2 \setminus \{Q\}} e^{-nI_{\tilde{Q}}(X;X')} \quad \text{(I.11)}$$

$$\doteq e^{n(R - I_Q(X;X'))} + \sum_{\tilde{Q} \in \mathcal{A}_1 \cup \mathcal{A}_2} e^{-nI_{\tilde{Q}}(X;X')} \quad \text{(I.12)}$$

$$\doteq e^{n(R - I_Q(X;X'))} + \max_{\tilde{Q} \in \mathcal{A}_1 \cup \mathcal{A}_2} e^{-nI_{\tilde{Q}}(X;X')}, \quad \text{(I.13)}$$

and hence

$$\Omega = \max_{Q \in \mathcal{A}_1 \cup \mathcal{A}_2} \Omega_Q \doteq \max_{Q \in \mathcal{A}_1 \cup \mathcal{A}_2} e^{n(R - I_Q(X;X'))}. \quad \text{(I.14)}$$

Furthermore,

$$\Phi \doteq \max_{Q \in \mathcal{A}_1 \cup \mathcal{A}_2} e^{-nI_Q(X;X')}, \quad \text{(I.15)}$$

such that

$$\Omega + \Phi \doteq \max_{Q \in \mathcal{A}_1 \cup \mathcal{A}_2} e^{n(R - I_Q(X;X'))} \tag{I.16}$$

$$= \max_{Q \in \mathcal{A}_2} e^{n(R - I_Q(X;X'))}. \tag{I.17}$$

Now, we would like to have $\Omega + \Phi \in [0, e^{-1}]$. Specifically, if $\Omega + \Phi \to 0$ as $n \to \infty$, then $\varphi(\Omega + \Phi) \doteq 1$. In order to have $\Omega + \Phi \to 0$, we need that

$$\max_{Q \in \mathcal{A}_2} \{R - I_Q(X;X')\} < 0, \tag{I.18}$$

or

$$\min_{Q_{XX'} \in \mathcal{A}_2} I_Q(X;X') > R. \tag{I.19}$$

Let us abbreviate $I_Q(X;X')$ by I_Q. In order to find the highest E_0 for which (I.19) holds, let us derive $\min_{Q_{XX'} \in \mathcal{A}_2} I_Q(X;X')$ as follows:

$$\min_{Q_{XX'} \in \mathcal{A}_2} I_Q$$

$$= \min_{\{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q \le 2R,\ \Gamma(Q, R-\epsilon) + I_Q - R \le E_0\}} I_Q \tag{I.20}$$

$$= \min_{Q_{XX'} \in \mathcal{Q}(Q_X)} \sup_{\sigma \ge 0} \sup_{\mu \ge 0} \{I_Q + \sigma \cdot (I_Q - 2R) + \mu \cdot (\Gamma(Q, R-\epsilon) + I_Q - R - E_0)\}, \tag{I.21}$$

where in (I.21) we used twice the fact that $\min_{\{Q:\ g(Q) \le 0\}} f(Q) = \min_Q \sup_{\sigma \ge 0} \{f(Q) + \sigma \cdot g(Q)\}$. For (I.21) to be strictly larger than R, it is equivalent to require that for all $Q_{XX'} \in \mathcal{A}_2$ there exist $\sigma \ge 0$ and $\mu \ge 0$ such that

$$I_Q + \sigma \cdot (I_Q - 2R) + \mu \cdot (\Gamma(Q, R-\epsilon) + I_Q - R - E_0) > R, \tag{I.22}$$

or, equivalently,

$$E_0 < \frac{I_Q - R + \sigma \cdot (I_Q - 2R)}{\mu} + \Gamma(Q, R-\epsilon) + I_Q - R. \tag{I.23}$$

Thus,

$$E_0 < \min_{Q_{XX'} \in \mathcal{Q}(Q_X)} \sup_{\mu \ge 0} \sup_{\sigma \ge 0} \left\{\Gamma(Q, R-\epsilon) + I_Q - R + \frac{I_Q - R + \sigma \cdot (I_Q - 2R)}{\mu}\right\} \tag{I.24}$$

$$= \min_{Q_{XX'} \in \mathcal{Q}(Q_X)} \sup_{\mu \ge 0} \left[\Gamma(Q, R-\epsilon) + I_Q - R + \begin{cases} \frac{I_Q - R}{\mu} & I_Q \le 2R \\ \infty & I_Q > 2R \end{cases}\right] \tag{I.25}$$

$$= \min_{\{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q \le 2R\}} \sup_{\mu \ge 0} \left\{\Gamma(Q, R-\epsilon) + I_Q - R + \frac{I_Q - R}{\mu}\right\} \tag{I.26}$$

$$= \min_{\{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q \le 2R\}} \left[\Gamma(Q, R-\epsilon) + I_Q - R + \begin{cases} 0 & I_Q \le R \\ \infty & I_Q > R \end{cases}\right] \tag{I.27}$$

$$= \min_{\{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q \le 2R,\ I_Q \le R\}} \{\Gamma(Q, R-\epsilon) + I_Q - R\} \tag{I.28}$$

$$= \min_{\{Q_{XX'} \in \mathcal{Q}(Q_X):\ I_Q \le R\}} \{\Gamma(Q, R-\epsilon) + I_Q - R\} \tag{I.29}$$

$$\equiv E_{\text{ex}}(R, \epsilon), \tag{I.30}$$

where the ∞ in (I.25) is because the maximizing $\sigma \ge 0$ in (I.24) when $I_Q > 2R$ is $\sigma^* = \infty$. The ∞ in (I.27) is due to the fact that when $I_Q > R$, the maximizing $\mu \ge 0$ in (I.26) is $\mu^* = 0$. Note that the exponent function $E_{\text{ex}}(R, \epsilon)$ converges to $E_{\text{ex}}(R)$ when $\epsilon \downarrow 0$. Finally, we use these results in Fact 3 and get the desired lower bound on $\mathbb{P}\{\mathcal{F}_0\}$.

REFERENCES

[1] A. Feinstein, "Error bounds in noisy channels without memory," *IEEE Trans. Inf. Theory*, vol. 1, no. 2, pp. 13–14, Sep. 1955.

[2] P. Elias, "Coding for noisy channels," *IRE Conv. Rec.*, vol. 4, pp. 37–46, 1955.

[3] R. M. Fano, *Transmission of Information*. Cambridge, MA, USA: MIT Press, 1961.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] N. Merhav, "Error exponents of typical random trellis codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2067–2077, Apr. 2020.

[10] N. Merhav, "A Lagrange–dual lower bound to the error exponent of the typical random code," *IEEE Trans. Inf. Theory*, early access, Dec. 31, 2019, doi: 10.1109/TIT.2019.2963392.

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

[12] 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.

[13] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.

[14] J. Liu, P. Cuff, and S. Verdú, "On α-decodability and α-likelihood decoder," in *Proc. 55th Ann. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Oct. 2017, pp. 118–124.

[15] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. Urbanke, "Comparing the bit-MAP and block-MAP decoding thresholds of reed-muller codes on BMS channels," 2016, *arXiv:1601.06048*. [Online]. Available: http://arxiv.org/abs/1601.06048

[16] A. Somekh-Baruch and N. Merhav, "Exact random coding exponents for erasure decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6444–6454, Oct. 2011.

[17] N. Merhav, "Exact random coding error exponents of optimal bin index decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6024–6031, Oct. 2014.

[18] N. Weinberger and N. Merhav, "Channel detection in coded communication," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6364–6392, Oct. 2017.

[19] S. Janson, "New versions of Suen's correlation inequality," *Random Struct. Algorithms*, vol. 13, nos. 3–4, pp. 467–483, Oct. 1998.

[20] A. Nazari, R. Venkataramanan, D. Krithivasan, S. S. Pradhan, and A. Anastasopoulos, "Typicality graphs: Large deviation analysis," 2010, *arXiv:1010.1317*. [Online]. Available: http://arxiv.org/abs/1010.1317

[21] N. Alon and J. Spencer, *The Probabilistic Method*. New York, NY, USA: Wiley, 1992.

[22] N. Merhav, "Statistical physics and information theory," *Found. Trends Commun. Inf.*, vol. 6, nos. 1–2, pp. 1–212, 2009.

Ran Tamir (Averbuch) (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 2010 and 2017, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering.

His research interests are in the areas of information theory, large deviations theory, and coded communications.

Neri Merhav (Fellow, IEEE) was born in Haifa, Israel, in March 1957. He received the B.Sc., M.Sc., and D.Sc. degrees from the Technion–Israel Institute of Technology, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1988 to 1990, he was with AT&T Bell Laboratories, Murray Hill, NJ, USA. Since 1990, he has been with the Electrical Engineering Department, Technion, where he is currently an Irving Shepard Professor. From 1994 to 2000, he served as a consultant to the Hewlett–Packard Laboratories–Israel (HPL-I). His research interests include information theory, statistical communications, and statistical signal processing. He is also especially interested in the areas of lossless/lossy source coding and prediction/filtering, relationships between information theory and statistics, detection, estimation, as well as in the area of Shannon theory, including topics in joint source–channel coding, source/channel simulation, and coding with side information with applications to information hiding and watermarking systems. Another recent research interest concerns the relationships between information theory and statistical physics.

Dr. Merhav was a co-recipient of the 1993 Paper Award of the IEEE Information Theory Society. He also received the 1994 American Technion Society Award for Academic Excellence and the 2002 Technion Henry Taub Prize for Excellence in Research. From 1996 until 1999, he has served as an Associate Editor for Source Coding to the IEEE TRANSACTIONS ON INFORMATION THEORY, where he has been an Associate Editor for Shannon Theory, since 2017. He has also served as a Co-Chairman of the Program Committee of the 2001 IEEE International Symposium on Information Theory. He is currently on the Editorial Board of FOUNDATIONS AND TRENDS IN COMMUNICATIONS AND INFORMATION THEORY.

Nir Weinberger (Member, IEEE) received the B.Sc. and M.Sc. degrees (both *summa cum laude)* from Tel Aviv University, Tel-Aviv, Israel, in 2006 and 2009, respectively, and the Ph.D. degree from the Technion–Israel Institute of Technology in 2017.

From 2017 to 2018, he was a Post-Doctoral Fellow with Tel Aviv University. From 2006 to 2013, he has served as an Algorithm Engineer with the Israeli Defense Forces, working in communications and signal processing. He is currently a Post-Doctoral Fellow with the Massachusetts Institute of Technology. He has been granted several scholarships for his Ph.D. studies and the 2017 Advanced Communication Center Award for his Ph.D. thesis.

Albert Guillén i Fàbregas (Senior Member, IEEE) received the degree in telecommunication engineering from the Universitat Politècnica de Catalunya and the degree in electronics engineering from the Politecnico di Torino, in 1999, and the Ph.D. degree in communication systems from the École Polytechnique Fédérale de Lausanne (EPFL) in 2004.

Since 2011, he has been an ICREA Research Professor with Universitat Pompeu Fabra. He is currently an Adjunct Researcher with the University of Cambridge. He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, the University of South Australia, the University of Cambridge, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, the University of South Australia, Centrum Wiskunde and Informatica, and Texas A&M University, Qatar. His specific research interests are in the areas of information theory, communication theory, coding theory, digital modulation, and signal processing techniques.

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 was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is also an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, and an Editor of the FOUNDATIONS AND TRENDS IN COMMUNICATIONS AND INFORMATION THEORY (Now Publishers).