Large Deviations of Typical Random Codes

Ran Averbuch and Neri Merhav
The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion – Israel Institute of Technology
rans@campus.technion.ac.il, merhav@ee.technion.ac.il

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

***Abstract*— This work contains two main contributions concerning the large deviations behavior of randomly chosen fixed composition codes over a discrete memoryless channel (DMC). The first is an exponentially tight expression for the probability of randomly drawing a codebook that performs worse than the typical random coding (TRC) error exponent, which is proved to be exponentially small. The second is lower and upper bounds on the probability of randomly selecting a codebook that outperforms the TRC error exponent, which turn out to be double–exponentially small, suggesting that relatively good codebooks are extremely rare. The key ingredient in the proofs is a new large deviations result of type class enumerators with dependent variables.**

I. Introduction

The first to show that below capacity, the probability of error decays exponentially with the block length, for a sequence of good codes, was Feinstein [1] in 1955. Already in the same year, Elias [2] derived the random coding bound and the sphere–packing bound, and he observed that they exponentially coincide at high rates, for the cases of the binary symmetric channel (BSC) and the binary erasure channel (BEC). Six years later, Fano [3] derived the random coding exponent, which is the limit of the negative normalized logarithm of the expectation (with respect to the ensemble of randomly generated codes) of the error probability, namely,

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

and heuristically also the sphere–packing bound for the general DMC. In 1965, Gallager [4] derived $E_{\rm r}(R)$ in a much simpler way and improved on $E_{\rm r}(R)$ at low rates by the idea of expurgation of randomly selected codes.

The typical random coding (TRC) error exponent is defined as the long–block limit of the negative normalized expectation of the logarithm of the error probability, i.e.,

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

We believe that the TRC exponent is the more relevant performance metric as it captures the most likely error exponent of a randomly selected code, 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. In addition, since in random coding analysis, the code is selected at random and remains fixed, it seems reasonable to study the performance of the very chosen code instead of directly considering the ensemble performance. Therefore, it is natural to ask what would be the error exponent associated with the typical randomly selected code.

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 as well as typical linear codes, for the special case of the BSC with maximum likelihood (ML) decoding. It was also shown that at a certain range of low rates, $E_{\rm trc}(R)$ lies between $E_{\rm r}(R)$ and the expurgated exponent, $E_{\rm ex}(R)$. In [6] Nazari *et al.* provided bounds on TRC error exponents for both DMCs and multiple–access channels. In a recent article by Merhav [7], an exact single–letter expression has been derived for the error exponent of typical, random, fixed 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.

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

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

where $P_{\rm e}(\mathcal{C}_n)$ is the error probability of a given code $\mathcal{C}_n$, governed by the randomness of the ensemble of codebooks $\{\mathcal{C}_n\}$. 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, as was shown in [7], while $\mathsf{E}(\mathcal{C}_n)$ concentrates at the TRC exponent, $E_{\rm trc}(R)$, here we are interested in probabilities of large fluctuations around $E_{\rm trc}(R)$.

Specifically, for a given $E_0 < E_{\rm trc}(R)$, we assess the probability of $\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) < E_0\right\}$ and provide an exponentially tight expression for it, which proves that bad codebooks are quite rare. In addition, we calculate the probability of $\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) > E_0\right\}$, where $E_0 > E_{\rm trc}(R)$, and derive a double–exponentially small lower and upper bounds on it. We find the largest value E_0, for which $\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) > E_0\right\}$ is strictly positive, by that proving the existence of exceptionally good codebooks.

The research of R. Averbuch and N. Merhav was supported by Israel Science Foundation (ISF) grant no. 137/18. The research of A. Guillén i Fàbregas has been funded in part by the European Research Council under ERC grant 725411 and by the Spanish Ministry of Economy and Competitiveness under grant TEC2016-78434-C3-1-R.

Due to the space limitation, technical details and proofs are omitted, but can be found in [9].

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 $\mathbb{E}[\cdot]$. 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, $\lim_{n\to\infty}(1/n)\log a_n/b_n = 0$, and similarly for $a_n \dot{\leq} b_n$ and $a_n \dot{\geq} b_n$. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathcal{I}\{\mathcal{E}\}$. The notation $[x]_+$ will stand for $\max\{0,x\}$.

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 $\hat{P}_{\boldsymbol{x}} = Q_X$.

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

$$W(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^{n} W(y_t|x_t). \tag{4}$$

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 fixed composition codes: for a given distribution Q_X, all of the vectors in $\mathcal{C}_n$ are uniformly and independently drawn from the type class $\mathcal{T}(Q_X)$. As in [7], [8], 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 $\boldsymbol{y}$:

$$\mathbb{P}\{\hat{m} = m|\boldsymbol{y}\} = \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}}{\sum_{m'=0}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}, \tag{5}$$

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 [10]). A more detailed discussion is given in [8].

Let $\boldsymbol{Y} \in \mathcal{Y}^n$ be the random channel output resulting from the transmission of $\boldsymbol{x}_m$. For a given code $\mathcal{C}_n$, define the error probability as

$$P_{\text{e}}(\mathcal{C}_n) = \frac{1}{M}\sum_{m=0}^{M-1}\mathbb{P}\{\hat{m}(\boldsymbol{Y}) \neq m|m \text{ sent}\}, \tag{6}$$

where $\mathbb{P}\{\cdot\}$ designates the probability measure associated with the randomness of the channel output given its input, and the randomness of the stochastic decoder.

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

$$E_{\text{trc}}(R, Q_X) = \lim_{n\to\infty}\left\{-\tfrac{1}{n}\mathbb{E}\left[\log P_{\text{e}}(\mathcal{C}_n)\right]\right\}. \tag{7}$$

In order to present the main result of [7], we define first a few quantities. Define the set $\mathcal{Q}(Q_X) = \{Q_{X'|X} : Q_{X'} = Q_X\}$ and

$$\alpha(R, Q_X, Q_Y) = \sup_{\mathcal{S}(Q_X, Q_Y)}[g(Q_{\tilde{X}Y}) - I_Q(\tilde{X};Y)] + R, \tag{8}$$

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

$$\begin{aligned}\Gamma(Q_{XX'}, R) = \inf_{Q_{Y|XX'}}\{&D(Q_{Y|X}\|W|Q_X) + I_Q(X';Y|X)\\ &+ [\max\{g(Q_{XY}), \alpha(R, Q_X, Q_Y)\} - g(Q_{X'Y})]_+\},\end{aligned} \tag{9}$$

where $D(Q_{Y|X}\|W|Q_X)$ is the conditional divergence between $Q_{Y|X}$ and W, averaged by Q_X. Under the above defined quantities, the TRC error exponent is given by [7]

$$E_{\text{trc}}(R, Q_X) = \inf_{\mathcal{S}(R, Q_X)}[\Gamma(Q_{XX'}, R) + I_Q(X;X') - R], \tag{10}$$

where $\mathcal{S}(R, Q_X) = \{Q_{X'|X} \in \mathcal{Q}(Q_X) : I_Q(X;X') \leq 2R\}$. The exponent $E_{\text{trc}}(R, Q_X)$ is the exact value around which the random variable $\mathsf{E}(\mathcal{C}_n)$ concentrates, as was proved in [7]. The expurgated exponent $E_{\text{ex}}(R, Q_X)$, proved in [8], has exactly the same expression, but with the minimization constraint in (10) $I_Q(X;X') \leq 2R$ replaced by $I_Q(X;X') \leq R$.

We are interested in two LD quantities, i.e., the lower and the upper tails of the distribution of $\mathsf{E}(\mathcal{C}_n)$. The first one is the probability that $\mathsf{E}(\mathcal{C}_n)$ is smaller than a given value E_0, lower than the expected value $E_{\text{trc}}(R, Q_X)$:

$$\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) < E_0\right\}, \tag{11}$$

which is the probability of drawing a *bad* codebook. The second is the probability that $\mathsf{E}(\mathcal{C}_n)$ exceeds a given E_0 (higher than $E_{\text{trc}}(R, Q_X)$):

$$\mathbb{P}\left\{\mathsf{E}(\mathcal{C}_n) > E_0\right\}, \tag{12}$$

which is the probability of drawing a *good* codebook. Finding exact expressions for (11) and (12) appears to be difficult. In this paper, we propose an exponentially tight expression for (11) and lower and upper bounds on (12), in order to characterize the existence of bad or good codes, respectively.

IV. Main Results

A. The Lower Tail

More refined questions concerning the lower tail are as follows. Does the probability $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) < E_0\}$ tend to zero with a finite exponent in the entire range $[0, E_{\text{trc}}(R, Q_X))$? If not, what is the range of E_0 for which $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) < E_0\}$ decays faster than exponentially? Answers to these questions would shed light and improve our understanding concerning the behavior of the ensemble of random fixed–composition codes.

In order to present the error exponent of the lower tail, we define a few more quantities. Define the set

$$\mathcal{L}(R, E_0, Q_X) \triangleq \{Q_{X'|X} \in \mathcal{Q}(Q_X) : [2R - I_Q(X;X')]_+ \geq \Gamma(Q_{XX'}, R) + R - E_0\}, \quad (13)$$

and the error exponent

$$E_{\text{lt}}(R, E_0, Q_X) \triangleq \inf_{\mathcal{L}(R,E_0,Q_X)} [I_Q(X;X') - 2R]_+. \quad (14)$$

Our first result in this section is the following theorem.

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

$$\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) < E_0\} \doteq \exp\{-n \cdot E_{\text{lt}}(R, E_0, Q_X)\}. \quad (15)$$

Understanding the behavior of the exponent $E_{\text{lt}}(R, E_0, Q_X)$ is important in order to gain a deeper understanding of random fixed composition codes. The following proposition provides a partial characterization for the behavior of the lower tail exponent.

Proposition 1: The function $E_{\text{lt}}(R, E_0, Q_X)$ has the following properties:

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

$$E_0^{\min}(R) = \inf_{\mathcal{Q}(Q_X)} \{\Gamma(Q_{XX'}, R) - [2R - I_Q(X;X')]_+\} + R. \quad (16)$$

Proposition 1 answers the questions we raised above. First, it proves 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 fixed 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 $(E_0^{\min}(R), E_{\text{trc}}(R, Q_X))$ tends to zero with a finite exponent, but for $E_0 \in [0, E_0^{\min}(R))$, the probability of drawing a codebook with $\mathsf{E}(\mathcal{C}_n) < E_0$ converges to zero faster than exponentially; these codebooks are extremely rare.

We next describe the behavior of $E_0^{\min}(R)$. Denote by $Q^*_{X'|X}(R)$ the minimizer of (16) 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 clipping in (16) is active and $E_0^{\min}(R)$ is given by

$$E_0^{\min}(R) = \inf_{\{\mathcal{Q}(Q_X):\ 2R \leq I_Q(X;X')\}} \Gamma(Q_{XX'}, R) + R, \quad (17)$$

which is a monotonically increasing function. On the other hand, if $R \geq R^*$, then the clipping in (16) is inactive and $E_0^{\min}(R)$ is simply given by the TRC error exponent $E_{\text{trc}}(R, Q_X)$. Figure 1 illustrates the error exponents, as well as $E_0^{\min}(R)$, for the binary z–channel with crossover parameter 0.001, a symmetric input distribution, $Q_X = (\frac{1}{2}, \frac{1}{2})$, and a ML decoder. The highest transmission rate is $R \cong 0.68524$.

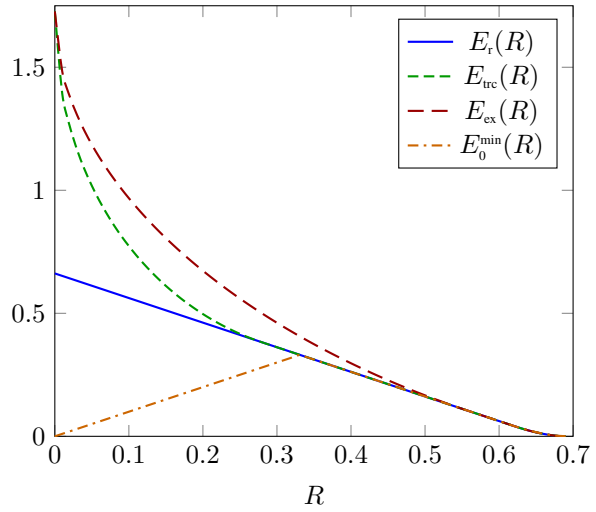


Fig. 1: Various exponents for the z–channel.

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

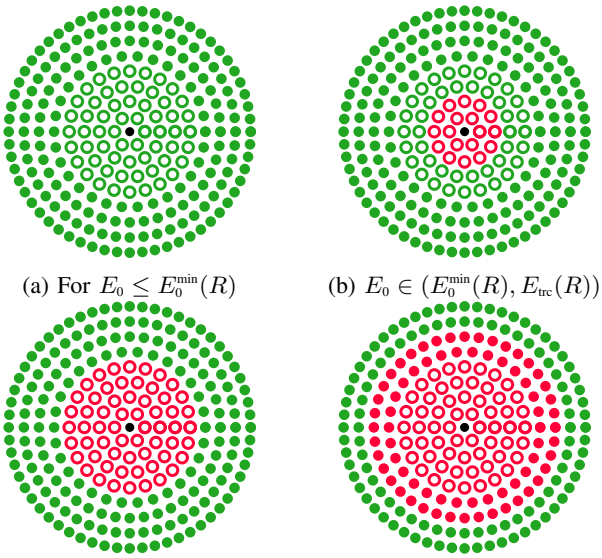
$$N(Q_{XX'}) \triangleq \sum_{m=0}^{M-1} \sum_{m' \neq m} \mathcal{I}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\}, \quad (18)$$

which play a pivotal role in the analysis of this lower tail exponent. The summation (18) contains $M(M-1) \doteq e^{n2R}$ terms. We refer to it as the *number of trials* in the type class enumerator $N(Q_{XX'})$. The expectation of each one of the binary random variables in (18) is given by $\mathbb{P}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\} \doteq e^{-nI_Q(X;X')}$, defined to be the *success probability*. It is important to note that $N(Q_{XX'})$ is not a binomial random variable.

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, those $Q_{XX'}$ for which $I_Q(X;X') > 2R$, are such that their corresponding $N(Q_{XX'})$ equals zero with high probability. These compositions are referred to as the *typically empty* (TE) type classes.

For $E_0 \in (E_0^{\min}(R), E_{\text{trc}}(R, Q_X))$, let us denote the minimizer of $E_{\text{lt}}(R, E_0, Q_X)$ by $Q^*_{X'|X}$ and define $Q^*_{XX'}$ by $Q_X \times Q^*_{X'|X}$. 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 Figure 2b). These are the joint type classes of the closest pairs of sequences in $\mathcal{X}^n$, as measured by the empirical mutual information.



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

(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 circles represent the TE type classes. TP type classes are the sets of relatively distant codewords; they include all joint compositions $Q_{XX'}$ with $I_Q(X;X') \leq 2R$. Red dots/circles mean empty type classes. For larger E_0 values, the minimum distance between codewords increases.

When E_0 is as low as the minimum $E_0^{\min}(R)$, the constraint set $\mathcal{L}(R, E_0, Q_X)$ becomes empty, all TE type classes become populated (see Figure 2a) and $E_{\text{lt}}(R, E_0, Q_X)$ jumps to infinity. In some sense, the curve $E_0^{\min}(R)$ exhibits a *phase transition*. When E_0 is higher than $E_0^{\min}(R)$, then the minimum distance between pairs of codewords is still positive, but when $E_0 \leq E_0^{\min}(R)$, the minimum distance between pairs of codewords equals zero.

For E_0 lower than $E_0^{\min}(R)$, we can explain the super–exponential behavior of $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) < E_0\}$ by the following intuitive example. If all TE type classes are populated, then codebooks with exponentially many identical codewords also exist in the range of these low exponents. Such poor codebooks are drawn with a double–exponentially small probability.

B. The Upper Tail

In this subsection, we discuss the existence of good codebooks for which $\mathsf{E}(\mathcal{C}_n) > E_{\text{trc}}(R, Q_X)$. Specifically, we study the probability of randomly drawing a codebook whose exponent $\mathsf{E}(\mathcal{C}_n)$ is higher than a given E_0, i.e., $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) > E_0\}$. Mainly, we are interested in lower–bounding the probability $\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) > E_0\}$, such that we can assure the existence of good codebooks. In order to present our results, we make the following definitions. Let

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

$$\Lambda(Q_{XX'}, R) = \inf_{Q_{Y|XX'}} \{D(Q_{Y|X}\|W|Q_X) + I_Q(X';Y|X) + \beta(R, Q_Y) - g(Q_{X'Y})\}. \quad (19)$$

We define the sets

$$\mathcal{V}(R, E_0, Q_X) = \{Q_{X'|X} \in \mathcal{Q}(Q_X) : \ I_Q(X;X') < 2R, \ \Lambda(Q_{XX'}, R) + I_Q(X;X') - R < E_0\}, \quad (20)$$

$$\mathcal{U}(R, E_0, Q_X) = \{Q_{X'|X} \in \mathcal{Q}(Q_X) : \ I_Q(X;X') < 2R, \ \Gamma(Q_{XX'}, R) + I_Q(X;X') - R < E_0\}, \quad (21)$$

and the error exponent functions

$$E_{\text{ut}}^{\text{ub}}(R, E_0, Q_X) = \sup_{\mathcal{V}(R,E_0,Q_X)} \min\{2R - I_Q(X;X'), \ E_0 - \Lambda(Q_{XX'}, R) - I_Q(X;X') + R, R\}, \quad (22)$$

$$E_{\text{ut}}^{\text{lb}}(R, E_0, Q_X) = \sup_{\mathcal{U}(R,E_0,Q_X)} [2R - I_Q(X;X')]. \quad (23)$$

The main result in this subsection is the following theorem.

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

$$\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) > E_0\} \dot{\leq} \exp\left\{-e^{n \cdot E_{\text{ut}}^{\text{ub}}(R,E_0,Q_X)}\right\}. \quad (24)$$

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

$$\mathbb{P}\{\mathsf{E}(\mathcal{C}_n) > E_0\} \dot{\geq} \exp\left\{-e^{n \cdot E_{\text{ut}}^{\text{lb}}(R,E_0,Q_X)}\right\}. \quad (25)$$

Since we have an upper, as well as a lower bound on the probability of the upper tail, both of which vanishes double–exponentially fast, we can state that (good) codebooks with relatively high error exponents are extremely rare.

The restriction to $(E_{\text{trc}}(R, Q_X), E_{\text{ex}}(R, Q_X))$ in the lower bound of Theorem 2 stems from the technical condition of [11, Theorem 9] (which is equivalent to the one found in the Lovász local lemma [12]).

In order to characterize the behavior of the error exponent functions (22) and (23), let us first define

$$\tilde{E}(R, Q_X) = \inf_{\mathcal{S}(R,Q_X)} [\Lambda(Q_{XX'}, R) + I_Q(X;X') - R]. \quad (26)$$

Proposition 2: $E_{\text{ut}}^{\text{ub}}(R, E_0, Q_X)$ and $E_{\text{ut}}^{\text{lb}}(R, E_0, Q_X)$ have the following properties:

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

Note that $\tilde{E}(R, Q_X)$ is defined similarly to $E_{\text{trc}}(R, Q_X)$, with $\Lambda(Q_{XX'}, R)$ replacing $\Gamma(Q_{XX'}, R)$. Generally, $\tilde{E}(R, Q_X) \geq E_{\text{trc}}(R, Q_X)$, but in some special cases, e.g. the z–channel and the BEC, it can be proved that $\tilde{E}(R, Q_X) = E_{\text{trc}}(R, Q_X)$, as can be seen in Figure 3 below.

Recall that for the typical code, i.e., any code with $\mathsf{E}(\mathcal{C}_n) \approx E_{\text{trc}}(R, Q_X)$, all TP type classes are populated and all TE type classes are empty (see Figure 2c). Now, for any E_0 in the range $(E_{\text{trc}}(R, Q_X), E_{\text{ex}}(R, Q_X))$, all TE type classes are still empty, but now, also all TP type classes that are associated with the set $\mathcal{U}(R, E_0, Q_X)$ are also empty (see Figure 2d). The dominating 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 (23). We conclude that $E_{\text{trc}}(R, Q_X)$ exhibits a second phase transition. Below the $E_{\text{trc}}(R, Q_X)$ curve, TE type classes become populated, and above it, TP type classes become empty.

When E_0 is as high as the expurgated exponent $E_{\text{ex}}(R, Q_X)$, the constraint set $\mathcal{U}(R, E_0, Q_X)$ contains all of the joint compositions with $R < I_Q(X; X') < 2R$, and all TP type classes associated with this range become empty. Furthermore, when E_0 reaches $E_{\text{ex}}(R, Q_X)$, then the value of the lower bound exponent $E^{\text{lb}}_{\text{ut}}(R, E_0, Q_X)$ is exactly R, which means that the lower bound of Theorem 2 and the probability of any codebook in the ensemble, given by $\exp\{-e^{nR}\}$, are of the same exponential order. We have the following corollary.

Corollary 1: If $E_0 < E_{\text{ex}}(R, Q_X)$, $\exists \mathcal{C}_n$ with $\mathsf{E}(\mathcal{C}_n) > E_0$.

Figure 3 illustrates the upper tail exponents, for the binary z–channel with crossover parameter 0.001, rate $R = 0.2$, a symmetric input distribution, $Q_X = (\frac{1}{2}, \frac{1}{2})$, and a ML decoder.

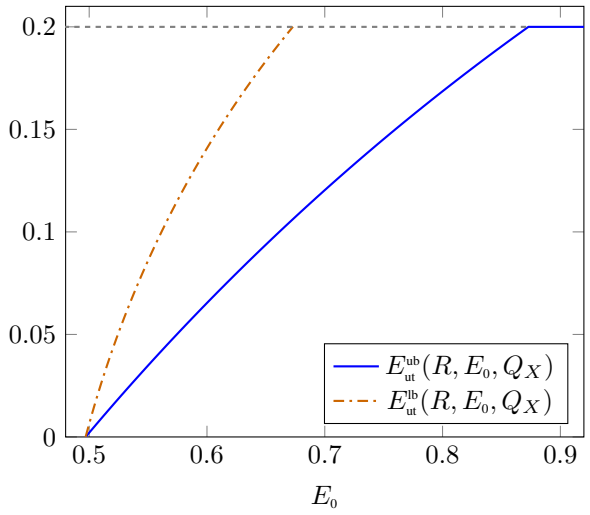


Fig. 3: Upper tail exponents for the z–channel ($R = 0.2$).

Note that $E^{\text{lb}}_{\text{ut}}(R, E_0, Q_X)$ is applicable as long as $0 \leq E^{\text{lb}}_{\text{ut}}(R, E_0, Q_X) \leq R$, while $E^{\text{ub}}_{\text{ut}}(R, E_0, Q_X)$ is applicable for any E_0, but is truncated to a level of R for relatively high E_0's.

V. On the Main Ingredient of the Proofs

The proofs of Theorems 1–2 are heavily based on LD analysis of the type class enumerators $N(Q_{XX'})$. One of the difficulties associated with the statistical analysis of $N(Q_{XX'})$ is that it is a sum of *dependent* (though pairwise independent) binary random variables. This is different from the more commonly encountered type class enumerators (see, e.g., [13, eq. (11)]), which are sums of *independent* random variables. Hence, existing results concerning the LD for type class enumerators of independent variables are no longer applicable.

Recall the following large–deviations property of ordinary type class enumerators. For some given sequence $\boldsymbol{y} \in \mathcal{Y}^n$, define the type class enumerator

$$N_{\boldsymbol{y}}(Q_{XY}) = \sum_{m=0}^{M-1} \mathcal{I}\{(\boldsymbol{x}_m, \boldsymbol{y}) \in \mathcal{T}(Q_{XY})\}. \tag{27}$$

Now, for any $t \in \mathbb{R}$, $\mathbb{P}\{N_{\boldsymbol{y}}(Q_{XY}) \geq e^{nt}\}$ tends to zero exponentially fast with $E(R, Q_{XY}, t)$, given by

$$\begin{cases} [I_Q(X;Y) - R]_+ & [R - I_Q(X;Y)]_+ \geq t \\ \infty & [R - I_Q(X;Y)]_+ < t \end{cases}. \tag{28}$$

Since $N(Q_{XX'})$ is a sum of dependent random variables, more refined tools from LD theory are required, like those of [11], that are able to treat the dependency between the terms. The LD of $N(Q_{XX'})$ and the ordinary type class enumerators asymptotically behave in the same way, as can be seen in the following result:

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

$$\mathbb{P}\{N(Q_{XX'}) \geq e^{ns}\} \doteq e^{-n \cdot E(R, Q_{XX'}, s)}, \tag{29}$$

where $E(R, Q_{XX'}, s)$ is given by

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

References

[1] A. Feinstein, "Error bounds in noisy channels without memory," *IRE Trans. Inf. Theory*, vol. IT–1, pp. 13–14, 1955.

[2] P. Elias, "Coding for noisy channels," *IRE Convention Record*, Part 4, pp. 37–46, 1955.

[3] R. M. Fano, *Transmission of Information*, MIT Press, Cambridge, Mass. and Wiley, New York 1961.

[4] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT–11, no. 1, pp. 3–18, 1965.

[5] A. Barg and G. D. Forney, Jr., "Random codes: minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, Sept. 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, Sept. 2014.

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

[8] 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.

[9] R. Averbuch, N. Merhav, and A. Guillén i Fàbregas, "Large Deviations of Typical Random Codes," in preparation.

[10] 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.

[11] S. Janson, "New versions of Suen's correlation inequality," *Random Structures Algorithms*, vol. 13, pp. 467–483, 1998.

[12] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, New York 1992.

[13] 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.