Concentration Properties of Generalized Random Gilbert-Varshamov Codes

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

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

***Abstract*—We study the typical error exponent of constant composition generalized random Gilbert-Varshamov (RGV) codes over discrete memoryless channels (DMC) channels with generalized likelihood decoding. We show that the typical error exponent of the RGV ensemble is equal to the expurgated error exponent, provided that the RGV codebook parameters are chosen appropriately. We also prove that the exponent of a randomly chosen RGV code converges in probability to the typical error exponent; the lower tail is shown to decay exponentially while the upper tail decays double-exponentially above the expurgated exponent.**

I. Introduction

Random coding is the key technique employed in information theory in order to show that a code with low error probability exists without explicitly constructing it. Codes are constructed at random, and the average error probability over all randomly generated codes is bounded. Then, it follows that there must exist a code with error probability at least as low as the ensemble average error probability over the codes. In particular, for discrete memoryless channel (DMC), Shannon [1] showed that there exists a code of rate smaller than the channel capacity with vanishing probability of error as the codeword length increases. For rates below capacity, Fano [2] characterized the exponential decay of the error probability defining the random coding exponent (RCE) as the negative normalized logarithm of the ensemble-average error probability. In [3], Gallager derived the RCE in a simpler way and introduced the idea of expurgation resulting in an improved exponent the at low rates.

Most proofs invoking random coding arguments, assume that codewords are independent. Random Gilbert-Varshamov (RGV) codes [4] are a family of random codes inspired by the basic code construction attaining the Gilbert-Varshamov bound in Hamming spaces. The code construction is based on drawing codewords recursively from a fixed type class, in such a way that a newly generated codeword must be at a certain minimum distance from all previously chosen codewords, according to some generic distance function. For suitably optimized distance functions, RGV codes attain Csiszár and Körner's (CK) exponent [5], which is known to be at least as high as both the random-coding and expurgated exponents.

In [6], Barg and Forney studied i.i.d. random coding over the binary symmetric channel (BSC) with maximum likelihood decoding and showed that the error exponent of most random codes is close to the so-called typical random coding (TRC) exponent, strictly larger than the RCE at low rates. Upper and lower bounds on the TRC for constant-composition codes and general DMCs were provided in [7]; Merhav [8] determined the exact TRC exponent with generalized likelihood decoders (GLD) for constant composition codes. Tamir *et al.* [9] studied the upper and lower tails of the error exponent around the TRC exponent for random pairwise-independent constant-composition codes with GLD. It was shown that the tails behave in a non-symmetric way: the lower tail decays exponentially while the upper tail decays doubly-exponentially. By studying the behavior of both tails, work in [9] proves concentration in probability. For pairwise-independent ensembles and arbitrary channels, Cocco *et al.* showed in [10] that the probability that a code in the ensemble has an exponent smaller than a lower bound on the TRC exponent is vanishingly small. Truong *et al.* showed that, for DMCs, the error exponent of a randomly generated code with pairwise-independent codewords converges in probability to its expectation – the TRC exponent [11].

This work focusses on the RGV code ensemble and discusses concentration properties of error exponents around its TRC. We find the exact typical error exponent TRC for the RGV ensemble by proving matched upper and lower bounds on the TRC and show it is equal to its RCE, i.e., to the maximum of the expurgated and random-coding exponent. We characterize the rates of the above convergence and show that it is exponential for the lower tail and double-exponential for the upper tail under some technical conditions. We show that the random error exponent converges in probability to the TRC. Compared with constant-composition codes with independent codewords, the dependence among RGV codewords causes standard concentration inequalities such as Hoeffding's inequality not to hold. In this work, we develop new techniques to overcome the challenges presented by RGV codeword dependence. Proofs of our results can be found in [12].

A. Notation

Random variables will be denoted by capital letters, and their realizations will be denoted by the corresponding lower

This work has been funded in part by the European Research Council under ERC grant agreement 725411 and by the Spanish Ministry of Economy and Competitiveness under grant PID2020-116683GB-C22.

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

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

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

A sequence of random variables $\{A_n\}_{n=1}^{\infty}$ is said to converge to A in probability, denoted as $A_n \xrightarrow{(\mathrm{p})} A$ if for all $\delta>0$ [13, Sec. 2.2],

$$\lim_{n\to\infty}\mathbb{P}[|A_n-A|>\delta]=0. \tag{2}$$

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

II. Preliminaries

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

$$W(\boldsymbol{y}|\boldsymbol{x}_m)=\prod_{i=1}^{n}W(y_i|x_{m,i}). \tag{3}$$

At the decoder side, we assume that a GLD [14] is used to infer what the transmitted message was. The GLD [14] extends the likelihood decoder in [15] and [16], and is a stochastic decoder that randomly selects the message estimate $\hat{m}$ according to the posterior probability distribution given the channel output $\boldsymbol{y}$ as follows

$$\Pr(\hat{m}=m|\boldsymbol{y})=\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_m,\boldsymbol{y}})\}}{\sum_{m=1}^{M}\exp\left\{ng(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})\right\}}, \tag{4}$$

where $g(\cdot)$, the *decoding metric*, is an arbitrary continuous function of a joint distribution P_{XY} on $\mathcal{X}\times\mathcal{Y}$. It can be shown that maximum likelihood and the well-known universal maximum mutual information (MMI) [17] decoders are particular instances of this GLD [14].

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

$$\begin{aligned}P_{\mathrm{e}}(\mathcal{C}_n)=\frac{1}{M}\sum_{m=1}^{M}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}W(\boldsymbol{y}|\boldsymbol{x}_m)\\ \times\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'},\boldsymbol{y}})\}}{\sum_{\tilde{m}=1}^{M}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}},\boldsymbol{y}})\}}.\end{aligned} \tag{5}$$

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

$$E_n(\mathcal{C}_n)=-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n). \tag{6}$$

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

III. RGV Codebook Ensemble and Properties

A. RGV Codebook Ensemble

In this section, we describe basic RGV codebook construction, channel model and GLD. The RGV codebook was first introduced in [4], which extended code constructions that attain the Gilbert-Varshamov bound on the Hamming space. The RGV construction is a randomized constant composition counterpart of such codes for arbitrary DMCs and arbitrary distance functions.

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

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

Hamming distance, Bhattacharyya distance, and equivocation distance [4].

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

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

$$\mathcal{T}(Q_X, \boldsymbol{x}_1) \triangleq \big\{\bar{\boldsymbol{x}} \in \mathcal{T}(Q_X) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_1) > \Delta\big\} \tag{7}$$

$$= \mathcal{T}(Q_X) \setminus \big\{\bar{\boldsymbol{x}} \in \mathcal{T}(Q_X) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_1) \leq \Delta\big\}, \tag{8}$$

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

$$\mathcal{T}(Q_X, \boldsymbol{x}_1^{i-1}) \triangleq \big\{\bar{\boldsymbol{x}} \in \mathcal{T}(Q_X) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_j) > \Delta, \\ j = 1, 2, \ldots, i-1\big\} \tag{9}$$

$$= \mathcal{T}(Q_X, \boldsymbol{x}_1^{i-2}) \setminus \big\{\bar{x} \in \mathcal{T}(Q_X, \boldsymbol{x}_1^{i-2}) : d(\bar{\boldsymbol{x}}, \boldsymbol{x}_{i-1}) \leq \Delta\big\} \tag{10}$$

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

For a given RGV code with rate R, type Q_X, distance function d, and minimum distance Δ and decoding metric g, we define the random coding exponent (RCE)

$$E_{\text{rce}}^{\text{rgv}}(R, Q_X, d, \Delta) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}[P_{\text{e}}(\mathcal{C}_n)] \tag{11}$$

and the typical random coding (TRC) error exponent as

$$E_{\text{trc}}^{\text{rgv}}(R, Q_X, d, \Delta) \triangleq \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\text{e}}(\mathcal{C}_n)], \tag{12}$$

provided that these limits exist, where the expectation is with respect to the randomness of the code $\mathcal{C}_n$.

Let $Q_X \in \mathcal{P}(\mathcal{X}), \Delta \in \mathbb{R}, d \in \Omega$, and define the following quantity

$$\Gamma(P_{XX'}, R) \triangleq \min_{P_{Y|XX'}} \Big\{D(P_{Y|X}\|W|Q_X) + I_P(X';Y|X) \\ + [\max\{g(P_{XY}), \alpha(R, P_Y)\} - g(P_{X'Y})]_+\Big\}, \tag{13}$$

where

$$\alpha(R, P_Y) \triangleq \max_{\substack{P_{X'|Y}: P_{X'} = Q_X, \\ I_P(X';Y) \leq R}} \big(g(P_{X'Y}) - I_P(X';Y)\big) + R. \tag{14}$$

The main result of [4] is that for ML decoding, and suitably optimized distance function and minimum distance, the constant composition RGV ensemble attains a random coding exponent equal to the CK exponent [5]. For GLD, we define the expurgated exponent as

$$E_{\text{ex}}(R, Q_X) \\ = \min_{\substack{P_{X'|X}: I_P(X;X') \leq R \\ P_{X'} = Q_X}} \{\Gamma(P_{XX'}, R) + I_P(X;X') - R\}. \tag{15}$$

is the expurgated exponent of the independent constant composition ensemble with composition Q_X and GLD. In this paper, we study the TRC of the RGV ensemble $E_{\text{trc}}^{\text{rgv}}(R, Q_X, d, \Delta)$ as well as the concentration of the exponent around the TRC. Specifically, we derive a generic expression of $E_{\text{trc}}^{\text{rgv}}(R, Q_X, d, \Delta)$ and show that $E_{\text{trc}}^{\text{rgv}}(R, Q_X, d, \Delta) = E_{\text{ex}}(R, Q_X)$ for suitably optimized minimum distance and distance functions. In addition, we provide bounds on the exponential and double-exponential concentration rates of the lower and upper tails of the error exponent of RGV codes.

IV. Main Results

In this section, we state some of our results. The first result is the typical error exponent.

Theorem 1: Let $Q_X \in \mathcal{P}(\mathcal{X}), \Delta \in \mathbb{R}, d \in \Omega$. Then, for any R satisfying

$$R \leq \min_{P_{XX'} \in \mathcal{Q}(Q_X): d(P_{XX'}) \leq \Delta} I(X;X') - 2\delta \tag{16}$$

for some $\delta > 0$, the typical random coding exponent of the RGV code ensemble with the GLD is given by

$$E_{\text{trc}}^{\text{rgv}}(R, Q_X, d, \Delta) \\ = \min_{\substack{P_{X'|X}: P_{X'} = Q_X, \\ I_P(X;X') \leq 2R, d(P_{XX'}) > \Delta}} \big\{\Gamma(P_{XX'}, R) + I_P(X;X') - R\big\}. \tag{17}$$

Similarly to [4], if we choose the distance function and minimum distance as $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -R$, respectively, we have that $E_{\text{trc}}^{\text{rgv}}(R, Q_X, d, \Delta) = E_{\text{ex}}(R, Q_X)$.

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

$$\mathcal{L}(R, E_0) \triangleq \{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \\ [2R - I_P(X;X')]_+ \geq \Gamma(P_{XX'}, R) + R - E_0\}, \tag{18}$$

$$\mathcal{M}(R, E_0) \triangleq \{P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \\ [2R - I_P(X;X')]_+ \geq \Lambda(P_{XX'}, R) + R - E_0\} \tag{19}$$

where

$$\Lambda(P_{XX'}, R) = \min_{P_{Y|XX'}} \big\{D(P_{Y|X}\|W|Q_X) + I_P(X';Y|X) \\ + \beta(R, P_Y) - g(P_{X'Y})\big\}, \tag{20}$$

$$\beta(R, P_Y) = \max_{P_{\tilde{X}|Y}: P_{\tilde{X}} = Q_X} \big\{g(P_{\tilde{X}Y}) + [R - I_P(\tilde{X};Y)]_+\big\}. \tag{21}$$

We have the following result.

Theorem 2: Consider the ensemble of RGV codes $\mathcal{C}_n$ of rate R and composition Q_X satisfying condition (16). Then, it holds that

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \leq E_0 \right] \dot{\leq} \exp\left\{ -nE_{\mathrm{lt}}^{\mathrm{ub}}(R,E_0) \right\}, \quad (22)$$

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \leq E_0 \right] \dot{\geq} \exp\left\{ -nE_{\mathrm{lt}}^{\mathrm{lb}}(R,E_0) \right\}. \quad (23)$$

where

$$E_{\mathrm{lt}}^{\mathrm{ub}}(R,E_0) \triangleq \min_{P_{XX'}\in\mathcal{L}(R,E_0)} [I_P(X;X') - 2R]_+, \quad (24)$$

$$E_{\mathrm{lt}}^{\mathrm{lb}}(R,E_0) \triangleq \min_{P_{XX'}\in\mathcal{M}(R,E_0)} [I_P(X;X') - 2R]_+, \quad (25)$$

respectively.

Next, we derive double-exponential upper and lower bounds to the upper tail probability. To begin with, we introduce additional notation. Let

$$\mathcal{V}(R,E_0) = \big\{ P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \\ I_P(X;X') \leq 2R, \Lambda(P_{XX'},R) + I_P(X;X') - R \leq E_0 \big\}, \quad (26)$$

$$\mathcal{U}(R,E_0) = \big\{ P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, \\ I_P(X;X') \leq 2R, \Gamma(P_{XX'},R) + I_P(X;X') - R \leq E_0 \big\}. \quad (27)$$

Define

$$\mathcal{A}_1 = \left\{ P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, I_P(X;X') > 2R \right\}, \quad (28)$$

$$\mathcal{A}_2 = \Big\{ P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, I_P(X;X') \leq 2R, \\ \Gamma(P_{XX'}, R-\varepsilon) + I_P(X;X') - R \leq E_0 + \varepsilon \Big\}, \quad (29)$$

and

$$\mathcal{A}_3 = \big\{ P_{XX'} \in \mathcal{Q}(Q_X) : d(P_{XX'}) > \Delta, I_P(X;X') \leq 2R, \\ \Gamma(P_{XX'}, R-\varepsilon) + I_P(X;X') - R > E_0 + \varepsilon \big\}. \quad (30)$$

Theorem 3: Consider the RGV ensemble $\mathcal{C}_n$ of rate R and composition Q_X satisfying condition (16). Then, the upper tail can be bounded as

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0 \right] \overset{\circ}{\leq} \exp\left\{ -\exp\left\{ nE_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0) \right\} \right\} \quad (31)$$

where

$$E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0) = \max_{P_{XX'}\in\mathcal{V}(R,E_0)} \min\big\{ 2R - I_P(X;X'), \\ E_0 - \Lambda(P_{XX'},R) - I_P(X;X') + R, R \big\}. \quad (32)$$

In addition, under the conditions

$$\max_{P_{XX'}\in\mathcal{A}_3} I_P(X;X') \leq \min_{P_{XX'}\in\mathcal{A}_2} I_P(X;X') \quad (33)$$

$$\min_{P_{XX'}:d(P_{XX'})\leq\Delta} I_P(X;X') \geq \max_{P_{XX'}:d(P_{XX'})>\Delta} I_P(X;X'), \quad (34)$$

and

$$\min_{P_{XX'}\in\mathcal{V}(R,E_0,\sigma)} I_P(X;X') - 2\delta \leq R \\ \leq \min_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})\leq\Delta} I_P(X;X') - 2\delta, \quad (35)$$

or

$$R \leq \min\Big\{ \min_{P_{XX'}\in\mathcal{V}}(R,E_0,\sigma) I_P(X;X') \\ - \min_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})\leq\Delta} I_P(X;X'), \\ \min_{P_{XX'}\in\mathcal{Q}(Q_X):d(P_{XX'})\leq\Delta} I_P(X;X') \Big\} - 2\delta \quad (36)$$

for some $\delta > 0$, we have that

$$\mathbb{P}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{C}_n) \geq E_0 \right] \overset{\circ}{\geq} \exp\left\{ -\exp\left\{ nE_{\mathrm{ut}}^{\mathrm{lb}}(R,E_0) \right\} \right\} \quad (37)$$

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

$$E_{\mathrm{ut}}^{\mathrm{lb}}(R,E_0) = \max_{P_{XX'}\in\mathcal{U}(R,E_0)} \big\{ 2R - I_P(X;X') \big\}. \quad (38)$$

V. Numerical Results

In Fig. 1, we plot various error exponents for the Z-channel with crossover probability 0.001 with $Q_X(0) = Q_X(1) = 1/2$. This example was considered in [9], [14]. Specifically, for reference we plot the random coding exponent $E_{\mathrm{r}}(R)$, the expurgated exponent $E_{\mathrm{ex}}(R)$, and the TRC $E_{\mathrm{trc}}(R)$ for constant composition codes. For the RGV ensemble exponents, we choose $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -R$ so as to achieve the largest possible exponents. We plot the corresponding random coding exponent $E_{\mathrm{rce}}^{\mathrm{rgv}}(R)$ and its corresponding TRC $E_{\mathrm{trc}}^{\mathrm{rgv}}(R)$ and illustrate that they both coincide with the expurgated exponent $E_{\mathrm{ex}}(R)$.

Fig. 2 shows exponential bounds for the lower tail for constant composition and the RGV ensemble with $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -R$. We observe that the lower and upper tails for RGV code ensemble decay faster than the lower and upper tails for the constant composition code. This can be explained by the the fact that, thanks to the structure of RGV codes, its error probability is expected to be smaller than that of constant composition codes.

In Figure 3 we show the double-exponential bounds for the upper tail for constant composition and the RGV ensemble with $d(P_{XX'}) = -I_P(X;X')$ and $\Delta = -R$. We observe that for constant composition the decay is indeed double-exponential even if the bounds only coincide for high values of E_0 (above the TRC exponent). Since $E_{\mathrm{ut}}^{\mathrm{ub}}(R,E_0)$ is double-exponential, we have that the upper tail decays at least double-exponentially above the expurgated exponent. The fact that

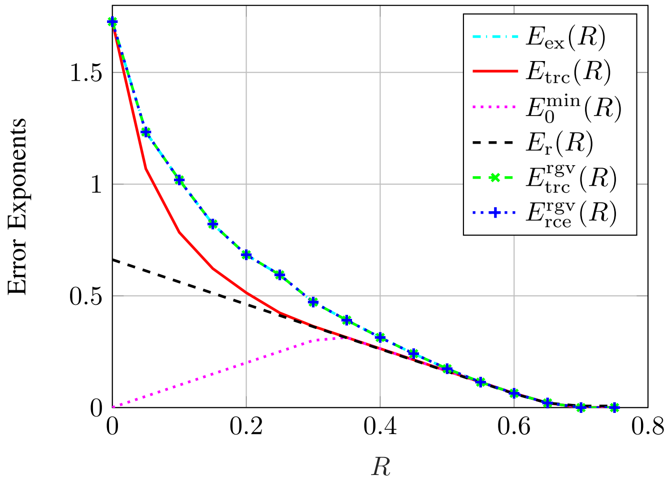

Fig. 1. Error Exponents for the Z-channel with crossover probability 0.001.

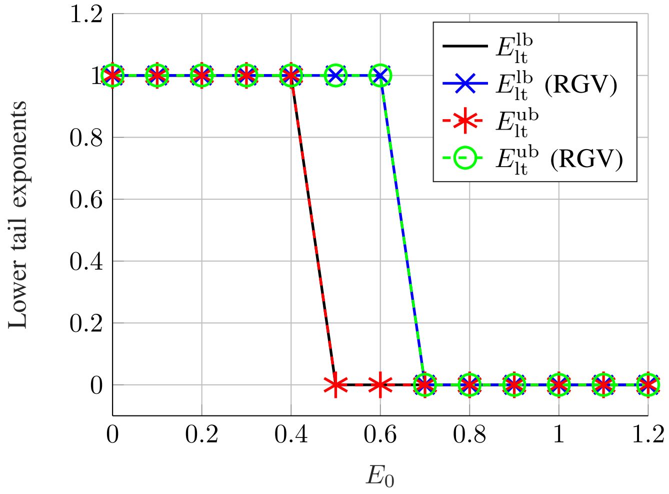

Fig. 2. Lower tail exponents for constant composition and RGV codes for the Z-channel with crossover probability 0.001.

$E_{\mathrm{ut}}^{\mathrm{lb}}(R, E_0) = 0$ below the expurgated exponent does not affect the double-exponential decay of the upper tail. Figure. 3 also shows that the decay rate of RGV code is slower than the constant composition code.

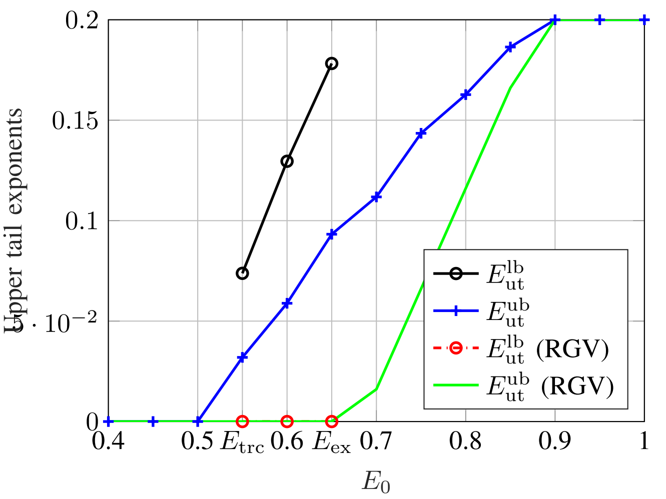

Fig. 3. Upper tail exponents for constant composition and RGV codes for the Z-channel with crossover probability 0.001.

References

[1] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

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

[3] R. G. Gallager. Simple derivation of the coding theorem and some applications. *IEEE Trans. Inf. Theory*, 11(3):3–18, Jan 2008.

[4] A. Somekh-Baruch, J. Scarlett, and A. Guillén i Fàbregas. Generalized random Gilbert-Varshamov codes. *IEEE Trans. Inf. Theory*, 65(6):3452–3469, 2019.

[5] I. Csiszár and J. Körner. Graph decomposition: A new key to coding theorems. *IEEE Trans. Inf. Th.*, 27:5–11, 1981.

[6] A. Barg and G. D. Forney. Random codes: minimum distances and error exponents. *IEEE Trans. Inf. Theory*, 48(9):2568–2573, 2002.

[7] A. Nazari, A. Anastasopoulos, and S. S. Pradhan. Error exponent for multiple-access channels: Lower bounds. *IEEE Trans. Inf. Theory*, 60(9):5095–5115, 2014.

[8] N. Merhav. Error exponents of typical random codes. *IEEE Trans. Inf. Theory*, 64(9):6223–6235, 2018.

[9] R. Tamir, N. Merhav, N. Weinberger, and A. Guillén i Fàbregas. Large deviations behavior of the logarithmic error probability of random codes. *IEEE Trans. Inf. Theory*, 66(11):6635–6659, 2020.

[10] G. Cocco, A. Guillén i Fàbregas, and J. Font-Segura. Typical error exponents: A dual domain derivation. *IEEE Trans. Inf. Theory*, to appear 2022.

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

[12] L. V. Truong and A. Guillén i Fàbregas. Generalized random Gilbert-Varshamov Codes: Typical error exponent and concentration properties. *ArXiv*, abs/2211.12238, 2022.

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

[14] N. Merhav. The generalized stochastic likelihood decoder: Random coding and expurgated bounds. *IEEE Trans. Inf. Th.*, 63(8):5039–5051, 2017.

[15] M. H. Yassaee, M. R. Aref, and A. Gohari. A technique for deriving one-shot achievability results in network information theory. In *2013 IEEE International Symposium on Information Theory*, pages 1287–1291, 2013.

[16] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. The likelihood decoder: Error exponents and mismatch. In *2015 IEEE Int. Symp. Inf. Theory*, pages 86–90, 2015.

[17] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.