IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Error Exponents of Typical Random Codes

Neri Merhav

CCIT Report #913
August 2017

Electronics
Computers
Communications

THE ANDREW & ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 3200003, ISRAEL

CCIT Report #913 August 2017

Error Exponents of Typical Random Codes

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

We define the error exponent of the typical random code as the long–block limit of the negative normalized *expectation of the logarithm* of the error probability of the random code, as opposed to the traditional random coding error exponent, which is the limit of the negative normalized *logarithm of the expectation* of the error probability. For the ensemble of uniformly randomly drawn fixed composition codes, we provide exact error exponents of typical random codes for a general discrete memoryless channel (DMC) and a wide class of (stochastic) decoders, collectively referred to as the generalized likelihood decoder (GLD). This ensemble of fixed composition codes is shown to be no worse than any other ensemble of independent codewords that are drawn under a permutation–invariant distribution (e.g., i.i.d. codewords). We also present relationships between the error exponent of the typical random code and the ordinary random coding error exponent, as well as the expurgated exponent for the GLD. Finally, we demonstrate that our analysis technique is applicable also to more general communication scenarios, such as list decoding (for fixed–size lists) as well as decoding with an erasure/list option in Forney's sense.

Index Terms: error exponent, typical random code, expurgated exponent, quenched average, likelihood decoder.

1 Introduction

Traditionally, the random coding error exponent is defined as

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

where n is the block length, R is the coding rate, $P_{\mathrm{e}}(\mathcal{C}_n)$ is the error probability of a codebook $\mathcal{C}_n$, and the expectation is with respect to (w.r.t) the randomness of $\mathcal{C}_n$ across the ensemble of codes (see, e.g., [2], [5], [14] and many references therein). While fairly easy to analyze (or at least, to bound), the random coding error exponent is also known to be quite a pessimistic performance measure because, at low coding rates, $\boldsymbol{E}P_{\mathrm{e}}(\mathcal{C}_n)$ is dominated by relatively poor codes in the ensemble, rather than by the channel noise. Indeed, at low coding rates, the random coding bound can be improved by the well known expurgation idea [2], [5], [14].

An alternative ensemble performance metric, that is never worse than the random coding error exponent, and in fact, strictly better at low rates, is the error exponent of the *typical random code* (TRC), which we define by simply commuting the expectation operator with the logarithmic function in (1), i.e.,

$$E_{\mathrm{trc}}(R)=\lim_{n\to\infty}\left[-\frac{\boldsymbol{E}\ln P_{\mathrm{e}}(\mathcal{C}_n)}{n}\right], \tag{2}$$

provided that the limit exists. The fact that $E_{\mathrm{trc}}(R)$ cannot be smaller than $E_{\mathrm{r}}(R)$ is easily understood from Jensen's inequality, but to capture the insight behind the different meanings of these two exponents, consider the following informal, intuitive consideration: let $\mathcal{S}(E)$ be the collection of all codes $\{\mathcal{C}_n\}$ in the ensemble, with $P_{\mathrm{e}}(\mathcal{C}_n)\approx e^{-nE}$ for a given value of E. Then, $\boldsymbol{E}P_{\mathrm{e}}(\mathcal{C}_n)\approx\sum_E \Pr\{\mathcal{S}(E)\}\cdot e^{-nE}$ (approximating by a discrete grid of values of E, for simplicity), a quantity that is dominated by the codes in $\mathcal{S}(E^*)$, where E^* maximizes the product $\Pr\{\mathcal{S}(E)\}\cdot e^{-nE}$. The codes of $\mathcal{S}(E^*)$ are the "poor" codes that we have referred to in the previous paragraph, and $E_{\mathrm{r}}(R)$ is given by E^* plus the exponential rate of $\Pr\{\mathcal{S}(E^*)\}$. On the other hand, $E_{\mathrm{trc}}(R)$ is approximately equal to $\sum_E \Pr\{\mathcal{S}(E)\}\cdot E$, and if there is one value of E, say E_0, at which $\Pr\{\mathcal{S}(E)\}$ concentrates in the large n limit, then the members of $\mathcal{S}(E_0)$ are the typical codes for our purpose, and $E_{\mathrm{trc}}(R)=E_0$. We will see later on that indeed, such a concentration property takes place, and hence the notion of "typical random codes". Generally speaking, we believe that the TRC error exponent should be the more relevant quantity of interest, because the code is selected randomly once and for all, and then it is natural to ask what would be the error exponent associated with the typical code.[1]

The problem is that it is considerably more difficult to analyze the expectation of the logarithm of the error probability than the logarithm of the expected error probability. This is true especially if one insists on obtaining exact error exponents and not just bounds. Perhaps this is one of the main reasons that not much earlier work has been done on error exponents of TRC's. The most relevant exception to this rule is the brief article by Barg and Forney [1] (see also [4]), where among other things, they have derived the error exponent of the TRC for the binary symmetric channel (BSC) w.r.t. the ensemble of codes drawn by fair coin tossing of each bit of each codeword. In [1], Barg and Forney have shown that at a certain range of low rates, $E_{\mathrm{trc}}(R)$ lies between $E_{\mathrm{r}}(R)$ and the expurgated exponent, $E_{\mathrm{ex}}(R)$, and there is an interesting relationship between $E_{\mathrm{trc}}(R)$ and the expurgated exponent function $E_{\mathrm{ex}}(\cdot)$ (also applicable to low rates), given by

$$E_{\mathrm{trc}}(R) = E_{\mathrm{ex}}(2R) + R. \tag{3}$$

Other related works, with some linkage to error exponents of TRC's, can be found in the statistical physics literature [6], [10], [12], where the replica method and the cavity method have been largely used mostly in the context of low–density parity–check (LDPC) codes.

In this work, we propose a systematic derivation of exact error exponents of TRC's. This extends the corresponding results of [1] in several directions.

1. A general DMC is considered, not merely the BSC.
2. The analysis covers a wide family of stochastic decoders, not only the maximum likelihood (ML) decoder (but the ML decoder is a special case).
3. We adopt the ensemble of constant composition codes, with independent codewords drawn under the uniform distribution across a given type class. This random coding distribution is shown to be no worse than any other permutation–invariant distribution, including, of course, the i.i.d. distribution, as in [1].

[1]Interestingly, there is an analogous consideration in statistical mechanics of disordered systems, which are modeled with random parameters. According to these models, Nature "selects" those random parameters just once, and one is interested in the free energy of a typical realization of the system, which is given in terms of the expectation of the logarithm of partition function (a.k.a. quenched average), rather than the logarithm of the expectation of the partition function (annealed average), see, e.g., [7, Sect. 5.7].

4. It is shown that the relation (3) continues to hold even in the more general scenario, as described in items 1–3 above. Moreover, using the improved expurgated exponent of [8], it is shown that eq. (3) holds for the entire range of rates, not merely at low rates.

5. It is demonstrated that the proposed analysis technique of TRC error exponents is applicable also to more general scenarios, such as list decoding (with fixed list size) as well as decoding with an erasure/list option in Forney's sense [3].

It should be pointed out that in [1, p. 2572, right column, comment no. 6], Barg and Forney comment that it is possible to extend the derivation to general DMC's with the ensemble of constant composition codes, but they have not displayed this extension, and it is not trivial to guess, from their analysis for the BSC and i.i.d. random coding, what is the TRC error exponent formula for a general DMC under the ensemble of constant composition codes and the more general decoders that we consider here.

The starting point of our analysis approach is similar to that of the well known replica method, a popular technique borrowed from statistical physics (see, e.g., [7, Sect. 4.5] and references therein), but this is the only point of similarity between our method and the replica method. In particular, it is based on the identity

$$\boldsymbol{E}\ln P_{\mathrm{e}}(\mathcal{C}_n) = \lim_{\rho\to\infty}\ln\left(\boldsymbol{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^{1/\rho}\right)^{\rho} = \lim_{\rho\to\infty}\rho\ln\left(\boldsymbol{E}[P_{\mathrm{e}}(\mathcal{C}_n)]^{1/\rho}\right). \tag{4}$$

This ingredient of calculating the $(1/\rho)$–th moment of the probability of error and raising it to the power of ρ, is also the technique used in the derivation of expurgated exponents [5], [14], the only difference is that in the context of expurgated bounds, this is applied to $P_{\mathrm{e}|m}(\mathcal{C}_n)$, the conditional error probability given that message m was transmitted, and then an expurgation argument is applied to assert that upon eliminating bad codewords from the code, we end up with a code for which $P_{\mathrm{e}|m}(\mathcal{C}_n)$ is upper bounded by a certain quantity, for all remaining messages. Here, on the other hand, we wish to invoke (4) for the overall error probability,

$$P_{\mathrm{e}}(\mathcal{C}_n) = \frac{1}{M}\sum_{m=0}^{M-1}P_{\mathrm{e}|m}(\mathcal{C}_n), \tag{5}$$

where M is the number of codebook messages. Nonetheless, this difference between the two derivations is not dramatic, and it is therefore not too surprising that the TRC error exponent and the expurgated exponent are related.

The outline of the remaining part of the paper is as follows. In Section 2, we establish notation conventions. In Section 3, we describe the setup, provide formal definitions, and spell out the objectives. In Section 4, we present the main result, a single–letter formula for the TRC error exponent, and discuss some of its properties. In Section 5, we prove the main result, and finally, in Section 6, we demonstrate how the same technique can be used to derive TRC error exponents in more general scenarios.

2 Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, ($n$ – positive integer) may take a specific vector value $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$, the n–th order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector. Sources and channels will be denoted by the letters P, Q and W, subscripted by the names of the relevant random variables/vectors and their conditionings, if applicable, following the standard notation conventions, e.g., Q_X, P_Y, $W_{Y|X}$, and so on. When there is no room for ambiguity, these subscripts will be omitted. The probability of an event $\mathcal{G}$ will be denoted by $\Pr\{\mathcal{G}\}$, and the expectation operator with respect to (w.r.t.) a probability distribution P will be denoted by $\boldsymbol{E}_P\{\cdot\}$. Again, the subscript will be omitted if the underlying probability distribution is clear from the context. 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, $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. Similarly, $a_n \overset{\cdot}{\le} b_n$ means that $\limsup_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} \le 0$, and so on. The indicator function of an event $\mathcal{G}$ will be denoted by $\mathcal{I}\{\mathcal{G}\}$. 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 type class of $\boldsymbol{x} \in \mathcal{X}^n$, denoted $\mathcal{T}(\hat{P}_{\boldsymbol{x}})$, is the set of all vectors $\boldsymbol{x}'$ with $\hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}$. Information measures associated with empirical distributions will be denoted with 'hats' and will be subscripted by the sequences from which they are induced. For example, the entropy associated with $\hat{P}_{\boldsymbol{x}}$, which is the empirical entropy of $\boldsymbol{x}$, will be denoted by $\hat{H}_{\boldsymbol{x}}(X)$. Similar conventions will apply to the joint empirical distribution, the joint type class, the conditional empirical distributions and the conditional type classes associated with pairs (and multiples) of sequences of length n. Accordingly, $\hat{P}_{\boldsymbol{x}\boldsymbol{y}}$ will be the joint empirical distribution of $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}_{i=1}^n$, and $\mathcal{T}(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})$ will denote the joint type class of $(\boldsymbol{x}, \boldsymbol{y})$. Similarly, $\mathcal{T}(\hat{P}_{\boldsymbol{x}|\boldsymbol{y}}|\boldsymbol{y})$ will stand for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{x}\boldsymbol{y}}(X, Y)$ will designate the empirical joint entropy of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{x}\boldsymbol{y}}(X|Y)$ will be the empirical conditional entropy, $\hat{I}_{\boldsymbol{x}\boldsymbol{y}}(X;Y)$ will denote empirical mutual information, and so on. We will also use similar rules of notation in the context of a generic distribution, Q_{XY} (or Q, for short, when there is no risk of ambiguity): we use $\mathcal{T}(Q_X)$ for the type class of sequences with empirical distribution Q_X, $H_Q(X)$ – for the corresponding empirical entropy, $\mathcal{T}(Q_{XY})$ – for the joint type class, $T(Q_{X|Y}|\boldsymbol{y})$ – for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $H_Q(X, Y)$ – for the joint empirical entropy, $H_Q(X|Y)$ – for the conditional empirical entropy of X given Y, $I_Q(X;Y)$ – for the empirical mutual information, and so on. We will also use the customary notation for the weighted divergence,

$$D(Q_{Y|X}\|P_{Y|X}|Q_X) = \sum_{x\in\mathcal{X}} Q_X(x) \sum_{y\in\mathcal{Y}} Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{P_{Y|X}(y|x)}. \tag{6}$$

Finally, the notation $(Q_Y \odot Q_{X|Y})_X$ will stand for the X–marginal induced by Q_Y and $Q_{X|Y}$, that is, $(Q_Y \odot Q_{X|Y})_X(x) = \sum_y Q_Y(y)Q_{X|Y}(x|y)$.

3 Formulation, Definitions, and Main Result

Consider a DMC, $W = \{W(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$, where $\mathcal{X}$ is a finite input alphabet, $\mathcal{Y}$ is a finite output alphabet, and $W(y|x)$ is the channel input–output single–letter transition probability from x to y. When fed by a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, the channel responds by producing an output vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$, according to

$$W(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1} W(y_i|x_i). \tag{7}$$

Let $\mathcal{C}_n = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\} \subseteq \mathcal{X}^n$, $M = e^{nR}$, R being 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 consider the ensemble of fixed composition codes, where each codeword is selected independently at random under the uniform distribution across a given type class of n–vectors, $\mathcal{T}(Q_X)$.

As in [8] and [9], we consider a generalized version of the so called *likelihood decoder* [11], [13], [15], which is a stochastic decoder that randomly selects the message estimate according to the posterior probability distribution given $\boldsymbol{y}$. The generalized likelihood decoder (GLD) considered here, randomly selects the decoded message according to the generalized posterior,

$$P(\hat{m}=m|\boldsymbol{y}) = \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}}{\sum_{m'=0}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}, \tag{8}$$

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

$$g(Q_{XY}) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} Q_{XY}(x,y)\ln W(y|x), \tag{9}$$

we recover the ordinary likelihood decoder as in [11], [13], [15]. For

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

$\beta\geq 0$ being a free parameter, we extend this to a parametric family of decoders, where β controls the skewedness of the posterior. In particular, $\beta\to\infty$ leads to the (deterministic) ML decoder. Other interesting choices are associated with mismatched metrics,

$$g(Q_{XY}) = \beta\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} Q_{XY}(x,y)\ln W'(y|x), \tag{11}$$

W' being different from W, and

$$g(Q_{XY}) = \beta I_Q(X;Y), \tag{12}$$

which for $\beta\to\infty$, approaches the well known universal maximum mutual information (MMI) decoder [2] (see also discussion around eqs. (5)–(7) of [8]). The probability of error, associated with a given code $\mathcal{C}_n$ and the GLD, is given by

$$P_{\mathrm{e}}(\mathcal{C}_n) = \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}_m)\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\sum_{\tilde{m}=0}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}. \tag{13}$$

For the ensemble of rate–R fixed composition codes of type Q_X, we define the random coding error exponent w.r.t. ML decoding (i.e., with g given by (10) at the limit $\beta\to\infty$), by

$$E_{\mathrm{r}}(R,Q_X) = \lim_{n\to\infty}\left[-\frac{\ln[\boldsymbol{E}P_{\mathrm{e}}(\mathcal{C}_n)]}{n}\right], \tag{14}$$

as well as the TRC error exponent, associated with the decoding metric g,

$$E_{\rm trc}^{g}(R,Q_X)=\lim_{n\to\infty}\left[-\frac{\boldsymbol{E}\ln[P_{\rm e}(\mathcal{C}_n)]}{n}\right], \tag{15}$$

provided that the limits exist,[2] and where the expectation is w.r.t. the randomness of $\mathcal{C}_n$. The TRC error exponent associated with the ML decoder will be denoted by $E_{\rm trc}(R,Q_X)$, and the one for the stochastic MMI decoder (that is, (12) with $\beta=1$) will be denoted by $E_{\rm trc}^{\rm smmi}(R,Q_X)$. The main objective of this paper is to derive an exact single–letter formula for $E_{\rm trc}^{g}(R,Q_X)$ and to study some of its properties.

4 Main Result

Before we present the main result, we need two more few definitions:

$$\alpha(R,Q_Y)\stackrel{\triangle}{=}\sup_{\{Q_{X|Y}:\ I_Q(X;Y)\le R,\ (Q_Y\odot Q_{X|Y})_X=Q_X\}}[g(Q_{XY})-I_Q(X;Y)]+R, \tag{16}$$

and

$$\begin{aligned}\Gamma(Q_{XX'},R)\ &\stackrel{\triangle}{=}\ \inf_{Q_{Y|XX'}}\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';Y|X)+\\&\qquad[\max\{g(Q_{XY}),\alpha(R,Q_Y)\}-g(Q_{X'Y})]_+\}.\end{aligned} \tag{17}$$

Our main result is the following theorem, whose proof appears in Section 5.

Theorem 1 *Consider the setting described in Section 3. Then,*

$$E_{\rm trc}^{g}(R,Q_X)=\inf_{\{Q_{X'|X}:\ I_Q(X;X')\le 2R,\ Q_{X'}=Q_X\}}\{\Gamma(Q_{XX'},R)+I_Q(X;X')-R\}. \tag{18}$$

The remaining part of this section is devoted to a discussion on Theorem 1 and its implications.

Relation to the random coding error exponent. In principle, the random coding error exponent is obtained by setting $\rho=1$ in the r.h.s. of (4) instead of taking the limit $\rho\to\infty$. We first show directly that $E_{\rm trc}(R,Q_X)$ indeed cannot be smaller than $E_{\rm r}(R,Q_X)$ at any rate R. Beyond the fact that this is a good sanity check, it is insightful to identify the origins of possible gaps

[2]The limit is well known to exist for (14). As for (15), it will be evident from the analysis.

between the two exponents. To this end, let us examine $E_{\text{trc}}^{\text{smmi}}(R,Q_X)$, that is, as mentioned before, defined for the sub-optimal GLD based on $g(Q)=I_Q(X;Y)$ (and which is especially convenient to work with). In this case, it can be readily verified that $\alpha(R,Q_Y)=R$, which yields

$$\Gamma(Q_{XX'})=\min_{Q_{Y|XX'}}\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';Y|X)+[\max\{I_Q(X;Y),R\}-I_Q(X';Y)]_+\}, \quad (19)$$

and so,

$$\begin{aligned}
E_{\text{trc}}(R,Q_X) &\geq E_{\text{trc}}^{\text{smmi}}(R,Q_X)\\
&= \min_{\{Q_{X'Y|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}}\Big\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';Y|X)+I_Q(X;X')+\\
&\quad [\max\{I_Q(X;Y),R\}-I_Q(X';Y)]_+-R\}\\
&= \min_{\{Q_{X'Y|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}}\Big\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';X|Y)+I_Q(X';Y)+\\
&\quad [\max\{I_Q(X;Y),R\}-I_Q(X';Y)]_+-R\}\\
&= \min_{\{Q_{X'Y|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}}\Big\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';X|Y)+\\
&\quad \max\{I_Q(X;Y),I_Q(X';Y),R\}-R\}\\
&= \min_{\{Q_{X'Y|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}}\Big\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';X|Y)+\\
&\quad [\max\{I_Q(X;Y),I_Q(X';Y)\}-R]_+\}\\
&\geq \min_{Q_{Y|X}}\Big\{D(Q_{Y|X}\|W|Q_X)+[I_Q(X;Y)-R]_+\Big\}\\
&= E_{\text{r}}(R,Q_X), \qquad (20)
\end{aligned}$$

where the first inequality is because the metric $g(Q)=I_Q(X;Y)$ may be sub-optimal and the second inequality is because we have dropped the constraints and the terms $I_Q(X';X|Y)$, and $I_Q(X';Y)$. The last equality is well known (see, e.g., [2]).

Relation to the expurgated exponent. It should be pointed out that in [8], the following expurgated bound was found for random fixed composition codes and the GLD:

$$E_{\text{ex}}^{g}(R,Q_X)=\inf_{\{Q_{XX'}:\ I_Q(X;X')\leq R,\ Q_{X'}=Q_X\}}\{\Gamma(Q_{XX'},R)+I_Q(X;X')-R\}, \quad (21)$$

and it has been shown in [8] that for ML decoding, this expurgated exponent is at least as large as the Csiszár–Körner–Marton (CKM) expurgated exponent [2, p. 165, Problem 10.18]. Obviously, we have the following simple relationship between $E_{\text{trc}}^{g}(R,Q_X)$ and $E_{\text{ex}}^{g}(R,Q_X)$:

$$E_{\text{trc}}^{g}(R,Q_X)=E_{\text{ex}}^{g}(2R,Q_X)+R, \quad (22)$$

which extends the relation (3) quite considerably. This relation is understood from the following consideration: as mentioned in the Introduction, the difference between the TRC error exponent and the expurgated exponent is that the former is applied to the overall error probability (5), whereas the latter is applied to the conditional error probability given that a particular message m was sent. The overall error probability (13) contains a double summation over the messages, indexed by m and m', whose exponential rate is $2R$, as opposed to the conditional error probability, which includes only a single summation over m', whose rate is R, hence the argument of $2R$ in the r.h.s. of (22). On the other hand, (13) contains normalization by M, which is absent in the conditional error probability, hence the addition of R on the r.h.s. of (22). Clearly, for any R, $E_{\mathrm{trc}}^g(R,Q_X) \le E_{\mathrm{ex}}^g(R)$, as the two functions are given by minimization of the same objective, but in $E_{\mathrm{trc}}^g(R,Q_X)$, the minimization is over a larger set of distributions. At zero–rate, we have $E_{\mathrm{trc}}^g(0,Q_X) = E_{\mathrm{ex}}^g(0,Q_X)$, which for ML decoding, is strictly larger than $E_{\mathrm{r}}(0,Q_X)$, in general. From continuity, it appears then that there is at least some range of low rates where the TRC error exponent is strictly larger than the random coding error exponent, but above a certain rate, the two exponents may coincide.

ML decoding. An important special case is, of course, the optimal ML decoder, which as mentioned earlier, corresponds to the choice $g(Q) = \beta E_Q \ln W(Y|X)$ for $\beta \to \infty$. For very large β, $\alpha(R,Q_Y) \approx \beta a(R,Q_Y)$, where

$$a(R,Q_Y) = \sup_{\{Q_{X|Y}:\ I_Q(X;Y)\le R,\ (Q_Y \odot Q_{X|Y})_X = Q_X\}} \boldsymbol{E}_Q \ln W(Y|X). \tag{23}$$

As $\beta \to \infty$, the term $\beta[\max\{E_Q \ln W(Y|X), a(R,Q_Y)\} - E_Q \ln W(Y|X')]_+$, that appears in the objective, disappears, and instead, there is an additional constraint that the expression in the square brackets of that term, would vanish. In other words, the result is

$$E_{\mathrm{trc}}(R,Q_X) = \inf_{Q_{X'Y|X}\in\mathcal{S}(R)} \{D(Q_{Y|X}\|W|Q_X) + I_Q(X';X,Y)\} - R, \tag{24}$$

where

$$\begin{aligned} \mathcal{S}(R) \ &\stackrel{\triangle}{=}\ \{Q_{X'Y|X}:\ I_Q(X;X') \le 2R,\ Q_{X'} = Q_X, \\ &\quad E_Q \ln W(Y|X') \ge \max\{E_Q \ln W(Y|X), a(R,Q_Y)\}\}. \end{aligned} \tag{25}$$

It is interesting to note that the third constraint in $\mathcal{S}(R)$ designates the event that an incorrect codeword (represented by X') receives a log–likelihood score higher than that of the correct codeword (represented by X) as well as those of all other codewords (represented by the term $a(R, Q)$). The term $a(R, Q)$ designates the typical value (with an extremely high probability) of the highest log–likelihood score among all the remaining incorrect codewords.[3] To understand the intuition behind this interpretation, observe that given a channel output $\boldsymbol{y} \in \mathcal{T}(Q_Y)$, the probability that a randomly chosen codeword from $\mathcal{T}(Q_X)$ would fall in a given conditional type, $\mathcal{T}(Q_{X|Y}|\boldsymbol{y})$, is of the exponential order of $e^{-nI_Q(X;Y)}$. Therefore, if we select e^{nR} codewords at random, all conditional types with $I_Q(X;Y) < R$ will be populated with very high probability. Among these conditional types, the highest log–likelihood score would be $\sup_{\{Q_{X|Y}:\ I_Q(X;Y) \le R,\ (Q_Y \odot Q_{X|Y})_X = Q_X\}} \boldsymbol{E}_Q \ln W(Y|X)$, which is exactly $a(R, Q_Y)$. This replaces the traditional union of *pairwise* error events, by the union of *disjoint* error events, where in each one of them, one incorrect codeword receives a score higher than all the others (not just higher than that of the correct codeword alone). As these events are disjoint, the probability of their union is equal to the sum of probabilities, i.e., the union bound is tight in this case.

Other ensembles with permutation–invariant random coding distributions. So far we considered only the ensemble of fixed composition codes, namely, each codeword was selected independently at random under the uniform distribution within $\mathcal{T}(Q_X)$. Consider, more generally, a probability distribution over $\mathcal{X}^n$ with the following two properties:

1. If $\boldsymbol{x}$ and $\boldsymbol{x}'$ belong to the same type, then $P(\boldsymbol{x}) = P(\boldsymbol{x}')$. In other words, the distribution is uniform within each type.

2. There exists a function, $\Delta(Q_X) \ge 0$, such that for every sequence, $\{Q_X^n\}$, of rational distributions with denominator n, and every Q_X at which $\Delta(Q_X)$ is continuous, $Q_X^n \to Q_X$ implies $\lim_{n\to\infty}[-\frac{1}{n} \log P\{\mathcal{T}(Q_X^n)\}] = \Delta(Q_X)$.

For example, if P is i.i.d., $\Delta(Q_X) = D(Q_X \| P)$. The ensemble of fixed composition codes also satisfies these requirements, provided that we allow some small tolerance δ in the empirical distribution rather than insisting on an exact empirical distribution,[4] and then

$$\Delta(Q_X) = \begin{cases} 0 & d(Q_X, Q_X^*) \le \delta \\ \infty & d(Q_X, Q_X^*) > \delta \end{cases} \tag{26}$$

[3]Observe that $a(R, Q)$ can be interpreted as the negative distortion–rate function (the inverse of the rate–distortion function) of a "source" Q_Y w.r.t. the distortion measure $d(x, y) = -\ln W(y|x)$ and the additional constraint that the "output" distribution would be Q_X.

[4]This small modification does not have any essential impact on the results.

where $d(\cdot,\cdot)$ is some distance measure in the space of distributions over $\mathcal{X}$.

It turns out, however, that there is nothing really to gain from this extension in terms of performance. In other words, among all ensembles of this family, the one of fixed composition codes, that we have studied thus far, is essentially the best. To see why this is true, consider the following argument, which is largely quite standard. Given a code $\mathcal{C}_n$, let $\mathcal{C}_n(Q_X) \stackrel{\triangle}{=} \mathcal{C}_n \cap \mathcal{T}(Q_X)$, $M(Q_X) \stackrel{\triangle}{=} |\mathcal{C}_n(Q_X)|$ and $R(Q_X) \stackrel{\triangle}{=} \frac{1}{n}\log M(Q_X)$. Obviously, there must be at least one Q_X for which $\Delta(Q_X) = 0$, since the number of different types is sub–exponential in n. Let us denote one of the distributions with this property by Q_X^*. Now, for every given $\mathcal{C}_n$, we have

$$\begin{aligned}
P_{\text{e}}(\mathcal{C}_n) &= \frac{1}{M}\sum_{m=0}^{M-1} P_{\text{e}|m}(\mathcal{C}_n) \\
&= \sum_{Q_X} \frac{M(Q_X)}{M} \cdot \frac{1}{M(Q_X)} \sum_{m:\ \boldsymbol{x}_m \in \mathcal{C}_n(Q_X)} P_{\text{e}|m}(\mathcal{C}_n) \\
&\geq \sum_{Q_X} \frac{M(Q_X)}{M} \cdot \frac{1}{M(Q_X)} \sum_{m:\ \boldsymbol{x}_m \in \mathcal{C}_n(Q_X)} P_{\text{e}|m}[\mathcal{C}_n(Q_X)] \\
&= \sum_{Q_X} \frac{M(Q_X)}{M} P_{\text{e}}[\mathcal{C}_n(Q_X)] \\
&\geq e^{n[R(Q_X)-R]} P_{\text{e}}[\mathcal{C}_n(Q_X)],
\end{aligned} \tag{27}$$

where the last inequality holds for every Q_X, and so,

$$\ln P_{\text{e}}(\mathcal{C}_n) \geq \ln P_{\text{e}}[\mathcal{C}_n(Q_X)] + n[R(Q_X) - R]. \tag{28}$$

Now, for every $\epsilon > 0$, as long as $\Delta(Q_X) < R$, with very high probability (tending to 1 double–exponentially rapidly w.r.t. the new ensemble), we will have $R(Q_X) \geq R - \Delta(Q_X) - \epsilon$, and in particular, $R(Q_X^*) \geq R - \Delta(Q_X^*) - \epsilon = R - \epsilon$, and so, for every such code

$$\ln P_{\text{e}}(\mathcal{C}_n) \geq \ln P_{\text{e}}[\mathcal{C}_n(Q_X^*)] - n\epsilon. \tag{29}$$

Let $\mathcal{G}_n$ denote the collection of codes with $R(Q_X) \geq R - \Delta(Q_X) - \epsilon$ for all Q_X such that $\Delta(Q_X) < R$, and observe that the probability of $\mathcal{G}_n$ is overwhelmingly large for large n. Then,

$$\begin{aligned}
\boldsymbol{E}\{\ln P_{\text{e}}(\mathcal{C}_n)\} &= \sum_{\mathcal{C}_n} P(\mathcal{C}_n)\ln P_{\text{e}}(\mathcal{C}_n) \\
&= \sum_{\mathcal{C}_n \in \mathcal{G}_n} P(\mathcal{C}_n)\ln P_{\text{e}}(\mathcal{C}_n) + \sum_{\mathcal{C}_n \in \mathcal{G}_n^c} P(\mathcal{C}_n)\ln P_{\text{e}}(\mathcal{C}_n)
\end{aligned}$$

$$\geq \quad P(\mathcal{G}_n) \cdot \sum_{\mathcal{C}_n \in \mathcal{G}_n} P(\mathcal{C}_n|\mathcal{G}_n) \ln P_{\mathrm{e}}[\mathcal{C}_n(Q_X^*)] - n\epsilon - n[E_{\mathrm{sp}}(R) + o(n)]P(\mathcal{G}_n^c)$$

$$\geq \quad -n[E_{\mathrm{trc}}^g(R-\epsilon, Q_X^*) + O(\epsilon)]P(\mathcal{G}_n) - n\epsilon - n[E_{\mathrm{sp}}(R) + o(n)]P(\mathcal{G}_n^c), \tag{30}$$

where we have used the fact that, due to the uniformity of the random coding distribution within each type, under $P(\cdot|\mathcal{G}_n)$, the sub-code $\mathcal{C}_n(Q_X^*)$ is a randomly selected fixed composition code of rate at least $R-\epsilon$. Now, since $P(\mathcal{G}_n^c)$ is double exponentially small, the right–most side is essentially $-nE_{\mathrm{trc}}^g(R, Q_X^*)$ for small ϵ.

5 Proof of Theorem 1

The proof of Theorem 1 is divided into two parts. In the first part, we prove that the TRC error exponent is lower bounded by the r.h.s. of eq. (18). In this proof, there are a few steps (such as the inequalities in eqs. (34) and (39) in the sequel) where it is not obvious that exponential tightness is not compromised, and therefore, we need the second part, where we prove that the TRC error exponent is also upper bounded by the same expression. Obviously, for the former, we need an upper bound on the error probability, whereas for the latter, we need a lower bound.

5.1 Lower Bound on the TRC Error Exponent

$$\begin{aligned}
P_{\mathrm{e}}(\mathcal{C}_n) &= \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\sum_{\tilde{m}=0}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}\\
&= \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}})\}+\sum_{\tilde{m}\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}\\
&\stackrel{\triangle}{=} \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}})\}+Z_m(\boldsymbol{y})},
\end{aligned} \tag{31}$$

and so, considering the ensemble of fixed composition codes of type Q_X, we have

$$\boldsymbol{E}\left\{[P_{\mathrm{e}}(\mathcal{C}_n)]^{1/\rho}\right\} = \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\cdot\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}})\}+Z_m(\boldsymbol{y})}\right]^{1/\rho}\right\}. \tag{32}$$

Let $\epsilon > 0$ be arbitrarily small. It is shown in [8, Appendix B], that with the possible exception of a double–exponentially small fraction of the fixed composition codes of type Q_X, all other codes in this class satisfy

$$Z_m(\boldsymbol{y}) \geq \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}, \qquad \forall\, m \in \{0, 1, \ldots, M-1\},\ \boldsymbol{y} \in \mathcal{Y}^n. \tag{33}$$

We then have,

$$\boldsymbol{E}\left\{[P_{\mathrm{e}}(\mathcal{C}_n)]^{1/\rho}\right\} \stackrel{\cdot}{\le} \boldsymbol{E}\left(\left[\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\ne m}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x}_m)\times \min\left\{1,\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{n\alpha(R-\epsilon,Q_Y)\}}\right\}\right]^{1/\rho}\right), \quad (34)$$

where we have neglected the double–exponentially small contribution of the codes that do not satisfy (33). Now, the inner–most sum (over $\{\boldsymbol{y}\}$) can be easily assessed using the method of types [2]. Using the arbitrariness of ϵ, the result[5] is that this sum is of the exponential order of $\exp\{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R)\}$, and so,

$$\begin{aligned}\boldsymbol{E}\left\{[P_{\mathrm{e}}(\mathcal{C}_n)]^{1/\rho}\right\} &\stackrel{\cdot}{\le} \boldsymbol{E}\left(\left[\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\ne m}\exp\{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R)\}\right]^{1/\rho}\right)\\ &= e^{-nR/\rho}\boldsymbol{E}\left\{\left[\sum_{Q_{XX'}}N(Q_{XX'})\exp\{-n\Gamma(Q_{XX'},R)\}\right]^{1/\rho}\right\}\\ &\stackrel{\cdot}{\le} e^{-nR/\rho}\sum_{Q_{XX'}}\boldsymbol{E}\{[N(Q_{XX'})]^{1/\rho}\}\cdot\exp\{-n\Gamma(Q_{XX'},R)/\rho\}, \end{aligned} \quad (35)$$

where $N(Q_{XX'})$ is the number of codeword pairs $\{(\boldsymbol{x}_m\boldsymbol{x}_{m'})\}$, $m'\ne m$, whose joint empirical distribution is $Q_{XX'}$, i.e.,

$$N(Q_{XX'})=\sum_m\sum_{m'\ne m}\mathcal{I}\{(\boldsymbol{x}_m\boldsymbol{x}_{m'})\in\mathcal{T}(Q_{XX'})\}, \quad (36)$$

and the summation is over all $\{Q_{XX'}\}$ whose marginals both agree with the given composition of the code, Q_X. We note that $N(Q_{XX'})$ can also be expressed as

$$N(Q_{XX'})=\sum_{m=0}^{M-1}N(Q_{X'|X}|\boldsymbol{x}_m), \quad (37)$$

where $N(Q_{X'|X}|\boldsymbol{x}_m)$ is the number of $\{\boldsymbol{x}_{m'}\}$, $m'\ne m$, that fall in the conditional type $\mathcal{T}(Q_{X'|X}|\boldsymbol{x}_m)$. Once we have an upper bound on $\boldsymbol{E}\{[N(Q_{XX'})]^{1/\rho}\}$, we can use it in order to bound

$$\boldsymbol{E}\ln P_{\mathrm{e}}(\mathcal{C}_n)\le\ln\left\{\sum_{Q_{XX'}}\lim_{\rho\to\infty}\left(\boldsymbol{E}\{[N(Q_{XX'})]^{1/\rho}\}\right)^{\rho}\cdot\exp\{-n\Gamma(Q_{XX'})\}\right\}-nR. \quad (38)$$

[5] See [8, Section V].

For a given $\rho > 1$, let $s \in [1, \rho]$. Then,

$$\begin{aligned}
\boldsymbol{E}\left\{[N(Q_{XX'})]^{1/\rho}\right\} &= \boldsymbol{E}\left[\sum_{m=0}^{M-1} N(Q_{X'|X}|\boldsymbol{x}_m)\right]^{1/\rho} \\
&= \boldsymbol{E}\left(\left[\sum_{m=0}^{M-1} N(Q_{X'|X}|\boldsymbol{x}_m)\right]^{1/s}\right)^{s/\rho} \\
&\leq \boldsymbol{E}\left(\sum_{m=0}^{M-1} \left[N(Q_{X'|X}|\boldsymbol{x}_m)\right]^{1/s}\right)^{s/\rho} \\
&\leq \left(\boldsymbol{E}\left\{\sum_{m=0}^{M-1} \left[N(Q_{X'|X}|\boldsymbol{x}_m)\right]^{1/s}\right\}\right)^{s/\rho} \\
&= \left(e^{nR}\boldsymbol{E}\left\{\left[N(Q_{X'|X}|\boldsymbol{x}_0)\right]^{1/s}\right\}\right)^{s/\rho} \\
&= e^{nRs/\rho}\left(\boldsymbol{E}\left\{\left[N(Q_{X'|X}|\boldsymbol{x}_0)\right]^{1/s}\right\}\right)^{s/\rho}, \qquad (39)
\end{aligned}$$

where the first inequality is based on the fact that $(\sum_i a_i)^t \leq \sum_i a_i^t$ whenever $\{a_i\}$ are non–negative and $t \in [0, 1]$, and the second inequality follows from the concavity of the function $f(u) = u^{s/\rho}$ when $0 < s/\rho \leq 1$. Now, for a given $\boldsymbol{x}_0$, $N(Q_{X'|X}|\boldsymbol{x}_0)$ is a binomial random variable with $e^{nR} - 1$ trials and success rate of the exponential order of $e^{-nI_Q(X;X')}$. Thus, similarly as shown in [7, eqs. (6.37), (6.38)],

$$\boldsymbol{E}\left\{\left[N(Q_{X'|X}|\boldsymbol{x}_0)\right]^{1/s}\right\} \doteq \begin{cases} \exp\{n[R - I_Q(X;X')]/s\} & R > I_Q(X;X') \\ \exp\{n[R - I_Q(X;X')]\} & R \leq I_Q(X;X') \end{cases} \qquad (40)$$

and so,

$$\boldsymbol{E}\left[N(Q_{XX'})\right]^{1/\rho} \stackrel{\cdot}{\leq} \begin{cases} \exp\{n[(s+1)R - I_Q(X;X')]/\rho\} & R > I_Q(X;X') \\ \exp\{ns[2R - I_Q(X;X')]/\rho\} & R \leq I_Q(X;X') \end{cases} \qquad (41)$$

which, after minimization over $s \in [1, \rho]$, becomes

$$\begin{aligned}
\boldsymbol{E}\left[N(Q_{XX'})\right]^{1/\rho} &\leq \min_{1\leq s\leq\rho} \begin{cases} \exp\{n[(s+1)R - I_Q(X;X')]/\rho\} & R > I_Q(X;X') \\ \exp\{ns[2R - I_Q(X;X')]/\rho\} & I_Q(X;X')/2 \leq R \leq I_Q(X;X') \\ \exp\{ns[2R - I_Q(X;X')]/\rho\} & R < I_Q(X;X')/2 \end{cases} \\
&= \begin{cases} \exp\{n[2R - I_Q(X;X')]/\rho\} & R > I_Q(X;X') \\ \exp\{n[2R - I_Q(X;X')]/\rho\} & I_Q(X;X')/2 \leq R \leq I_Q(X;X') \\ \exp\{n\rho[2R - I_Q(X;X')]/\rho\} & R < I_Q(X;X')/2 \end{cases} \\
&= \begin{cases} \exp\{n[2R - I_Q(X;X')]/\rho\} & R > I_Q(X;X')/2 \\ \exp\{n[2R - I_Q(X;X')]\} & R < I_Q(X;X')/2 \end{cases} \qquad (42)
\end{aligned}$$

and so,

$$\left(\boldsymbol{E}\left[N(Q_{XX'})\right]^{1/\rho}\right)^{\rho} \leq \begin{cases} \exp\{n[2R - I_Q(X;X')]\} & R > I_Q(X;X')/2 \\ \exp\{n\rho[2R - I_Q(X;X')]\} & R < I_Q(X;X')/2 \end{cases} \qquad (43)$$

which in the limit becomes

$$\lim_{\rho\to\infty}\left(\boldsymbol{E}\left[\sum_m N(Q_{X'|X}|\boldsymbol{x}_m)\right]^{1/\rho}\right)^\rho \le \begin{cases}\exp\{n[2R-I_Q(X;X')]\} & R > I_Q(X;X')/2\\ 0 & R < I_Q(X;X')/2\end{cases} \quad (44)$$

and substituting this into (38), yields

$$\begin{aligned} E^g_{\rm trc}(R,Q_X) &\ge \min_{\{Q_{XX'}:\ I_Q(X;X')\le 2R,\ Q_{X'}=Q_X\}}\{\Gamma(Q_{XX'})+I_Q(X;X')-2R+R\}\\ &= \min_{\{Q_{XX'}:\ I_Q(X;X')\le 2R,\ Q_{X'}=Q_X\}}\{\Gamma(Q_{XX'})+I_Q(X;X')-R\}\\ &= E^g_{\rm ex}(2R)+R, \end{aligned} \quad (45)$$

completing half of the proof of Theorem 1.

5.2 Upper Bound on the TRC Error Exponent

The idea of the proof is show that with high probability, the randomly selected code is such that $Z_{mm'}(\boldsymbol{y})$ is *upper* bounded by $\exp\{n[\alpha(R+2\epsilon,\hat{P}_{\boldsymbol{y}})+\epsilon]\}$ for sufficiently many triplets $\{(m,m',\boldsymbol{y})\}$, and so, the denominator of the generalized posterior can be lower bounded by an expression of the same exponential order as before.

We begin with a simple fact that will be needed later. Consider a joint distribution, $Q_{XX'}$, that satisfies $I_Q(X;X') < 2R$, and define $\mathcal{E}(Q_{XX'}) = \{\mathcal{C}_n:\ N(Q_{XX'}) < \exp\{n[2R-I_Q(X;X')-\epsilon\}\}$. We have to show that $\Pr\{\mathcal{E}(Q_{XX'})\}$ is small.[6] This follows from the following consideration.

$$\begin{aligned} \Pr\{\mathcal{E}(Q_{XX'})\} &= \Pr\{N(Q_{XX'}) < \exp\{n[2R-I_Q(X;X')-\epsilon\}\}\\ &= \Pr\{N(Q_{XX'}) < e^{-n\epsilon}\boldsymbol{E}\{N(Q_{XX'})\}\}\\ &= \Pr\left\{\frac{N(Q_{XX'})}{\boldsymbol{E}\{N(Q_{XX'})\}}-1 < -(1-e^{-n\epsilon})\right\}\\ &\le \Pr\left\{\left[\frac{N(Q_{XX'})}{\boldsymbol{E}\{N(Q_{XX'})\}}-1\right]^2 > (1-e^{-n\epsilon})^2\right\}\\ &\le \frac{\mathrm{Var}\{N(Q_{XX'})\}}{(1-e^{-n\epsilon})^2\boldsymbol{E}^2\{N(Q_{XX'})\}}. \end{aligned} \quad (46)$$

[6] This fact, together with the small probability of the event $\{\mathcal{C}_n:\ N(Q_{XX'}) > \exp\{n[2R-I_Q(X;X')+\epsilon\}\}$ (proved similarly) means that $N(Q_{XX'})$ concentrates around its mean. Also, when $I_Q(X;X') > 2R$, we have $\Pr\{N(Q_{XX'}) \ge 1\} \le \boldsymbol{E}\{N(Q_{XX'})\} \to 0$. Therefore, the typical codes (referring to $\mathcal{S}(E_0)$ in the Introduction), which capture most of the probability, are characterized by $N(Q_{XX'}) = 0$ for all $I_Q(X;X') > 2R$, and $N(Q_{XX'}) \doteq \exp\{n[2R-I_Q(X;X')]\}$ for all $I_Q(X;X') < 2R$.

Now, the denominator is of the exponential order of $\exp\{2n[2R - I_Q(X;X')]\}$. Using the shorthand notation $\mathcal{I}(m,m') = \mathcal{I}\{(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\}$ and $p = \boldsymbol{E}\mathcal{I}(m,m')$, the numerator is given as follows:

$$\begin{aligned} \text{Var}\{N(Q_{XX'})\} &= \boldsymbol{E}\{N^2(Q_{XX'})\} - \boldsymbol{E}^2\{N(Q_{XX'})\} \\ &= \sum_{m,m'}\sum_{\tilde{m},\hat{m}} \boldsymbol{E}\{\mathcal{I}(m,m')\mathcal{I}(\tilde{m},\hat{m})\} - [M(M-1)p]^2 \\ &= \sum_{m,m'} \boldsymbol{E}\{\mathcal{I}^2(m,m')\} + \sum_{(m,m')\neq(\tilde{m},\hat{m})} \boldsymbol{E}\{\mathcal{I}(m,m')\mathcal{I}(\tilde{m},\hat{m})\} - [M(M-1)p]^2 \\ &= M(M-1)p + M(M-1)[M(M-1)-1]p^2 - [M(M-1)p]^2 \\ &= M(M-1)p[1 + (M(M-1)-1)p - M(M-1)p] \\ &= M(M-1)p(1-p) \\ &\doteq \exp\{n[2R - I_Q(X;X')]\}. \end{aligned} \tag{47}$$

Thus,

$$\begin{aligned} \Pr\{\mathcal{E}(Q_{XX'})\} &\stackrel{\cdot}{\leq} \frac{\exp\{n[2R - I_Q(X;X')]\}}{\exp\{n[4R - 2I_Q(X;X')]\}} \\ &= \exp\{-n[2R - I_Q(X;X')]\}, \end{aligned} \tag{48}$$

which tends to zero since we have assumed that $I_Q(X;X') < 2R$. Of course, if $I_Q(X;X') < 2R - \epsilon$, then $\Pr\{\mathcal{E}(Q_{XX'})\}$ decays at least as fast as $e^{-n\epsilon}$.

Next, for a given $\epsilon > 0$, and a given joint type, $Q_{XX'Y}$, such that $I_Q(X;X') < 2R - \epsilon$, let us define

$$Z_{mm'}(\boldsymbol{y}) = \sum_{\tilde{m}\neq m,m'} \exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}, \tag{49}$$

and

$$\begin{aligned} \mathcal{G}_n(Q_{XX'Y}) = \Bigg\{\mathcal{C}_n : &\sum_m \sum_{m'\neq m} \mathcal{I}\{(\boldsymbol{x}_m \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\} \times \\ &\sum_{\boldsymbol{y}\in\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})} \mathcal{I}\{Z_{mm'}(\boldsymbol{y}) \leq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\} \geq \\ &\exp\{n[2R - I_Q(X;X') - 3\epsilon/2]\} \cdot |\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|\Bigg\}, \end{aligned} \tag{50}$$

where $(\boldsymbol{x}_m\boldsymbol{x}_{m'})$, in the expression $|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|$, should be understood as any pair of codewords in $\mathcal{T}(Q_{XX'})$ (as the specific choice of them is immaterial for the size of $\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})$).

Next define,

$$\mathcal{G}_n = \bigcap_{\{Q_{XX'Y}:\ I_Q(X;X')<2R-\epsilon\}} [\mathcal{G}_n(Q_{XX'Y}) \cap \mathcal{E}^c(Q_{XX'})]. \tag{51}$$

We first show that $\Pr\{\mathcal{G}_n\} \to 1$ as $n \to \infty$. We have already shown that $\Pr\{\mathcal{E}(Q_{XX'})\} \leq e^{-n[2R-I_Q(X;X')]} \leq e^{-n\epsilon}$. As for $\mathcal{G}_n(Q_{XX'Y})$, we have the following consideration. By the Chebychev inequality

$$\begin{aligned}
& \Pr\{[\mathcal{G}_n(Q_{XX'Y})]^c\} \\
\leq\ & \Pr\left[\sum_m \sum_{m'\neq m} \mathcal{I}\{(\boldsymbol{x}_m \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\} \cdot \sum_{\boldsymbol{y}\in\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})} \mathcal{I}\{Z_{mm'}(\boldsymbol{y}) > e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\} > \right. \\
& \left. \exp\{n[2R - I_Q(X;X') - 3\epsilon/2]\} \cdot |\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|\right] \\
\leq\ & \frac{\boldsymbol{E}\left\{\sum_m \sum_{m'\neq m} \mathcal{I}\{(\boldsymbol{x}_m\boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\} \cdot \sum_{\boldsymbol{y}\in\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})} \mathcal{I}\{Z_{mm'}(\boldsymbol{y}) > e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\}\right\}}{\exp\{n[2R - I_Q(X;X') - 3\epsilon/2]\} \cdot |\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|} \\
\leq\ & \frac{e^{2nR}|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})| \cdot \Pr\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'}),\ Z_{mm'}(\boldsymbol{y}) \geq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\}}{\exp\{n[2R - I_Q(X;X') - 3\epsilon/2]\} \cdot |\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|} \\
=\ & \frac{\Pr\{(\boldsymbol{X}_m, \boldsymbol{X}_{m'}) \in \mathcal{T}(Q_{XX'})\} \cdot \Pr\{Z_{mm'}(\boldsymbol{y}) \geq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\}}{\exp\{-n[I_Q(X;X') + 3\epsilon/2]\}} \\
\doteq\ & \frac{e^{-nI_Q(X;X')} \cdot \Pr\{Z_{mm'}(\boldsymbol{y}) \geq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\}}{\exp\{-n[I_Q(X;X') + 3\epsilon/2]\}} \\
=\ & e^{3n\epsilon/2} \cdot \Pr\{Z_{mm'}(\boldsymbol{y}) \geq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\}.
\end{aligned} \tag{52}$$

But

$$\begin{aligned}
& \Pr\{Z_{mm'}(\boldsymbol{y}) \geq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\} \\
=\ & \Pr\left\{\sum_{Q_{X|Y}} N(Q_{XY}) e^{ng(Q_{XY})} \geq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\right\} \\
\doteq\ & \max_{Q_{X|Y}} \Pr\left\{N(Q_{XY}) \geq \exp\{n[\alpha(R+2\epsilon,Q_Y) + \epsilon - g(Q_{XY})]\}\right\} \\
\doteq\ & e^{-nE},
\end{aligned} \tag{53}$$

where $N(Q_{XY})$ is the number of codewords other than $\boldsymbol{x}_m$ and $\boldsymbol{x}_{m'}$ that, together with $\boldsymbol{y}$, fall in $\mathcal{T}(Q_{XY})$, which is a binomial random variables with $e^{nR} - 2$ trials and success rate of exponential order $e^{-nI_Q(X;Y)}$, and so,

$$\begin{aligned}
E &= \min_{Q_{X|Y}} \begin{cases} [I_Q(X;Y) - R]_+ & g(Q_{XY}) + [R - I_Q(X;Y)]_+ \geq \alpha(R+2\epsilon, Q_Y) + \epsilon \\ \infty & g(Q_{XY}) + [R - I_Q(X;Y)]_+ < \alpha(R+2\epsilon, Q_Y) + \epsilon \end{cases} \\
&= \inf_{\{Q_{X|Y}:\ g(Q_{XY})+[R-I_Q(X;Y)]_+ \geq \alpha(R+2\epsilon,Q_Y)+\epsilon\}} [I_Q(X;Y) - R]_+.
\end{aligned} \tag{54}$$

Now, by definition of the function $\alpha(R, Q_Y)$, the set $\{Q_{X|Y}: \ g(Q_{XY}) + [R - I_Q(X;Y)]_+ \geq \alpha(R+2\epsilon, Q_Y) + \epsilon\}$ is a subset of $\{Q_{X|Y}: \ I_Q(X;Y) \geq R + 2\epsilon\}$. Thus,

$$E \geq \inf_{\{Q_{X|Y}: \ I_Q(X;Y) \geq R+2\epsilon\}} [I_Q(X;Y) - R]_+ = 2\epsilon, \tag{55}$$

and so, $\Pr\{Z_{mm'}(\boldsymbol{y}) \geq e^{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]}\} \stackrel{\cdot}{\leq} e^{-2n\epsilon}$, which leads to

$$\Pr\{[\mathcal{G}_n(Q_{XX'Y})]^c\} \stackrel{\cdot}{\leq} e^{3n\epsilon/2} \cdot e^{-2n\epsilon} = e^{-n\epsilon/2}. \tag{56}$$

Since the number of types is merely polynomial, it follows that $\Pr\{\mathcal{G}_n\} \to 1$. Now, for a given $\mathcal{C}_n \in \mathcal{G}(Q_{XX'Y})$, let us define the set

$$\mathcal{F}(\mathcal{C}_n, Q_{XX'Y}) = \{(m, m', \boldsymbol{y}): \ Z_{mm'}(\boldsymbol{y}) \leq \exp\{n[\alpha(R+2\epsilon, Q_Y) + \epsilon]\}\}, \tag{57}$$

and

$$\mathcal{F}(\mathcal{C}_n, Q_{XX'Y}|m, m') = \{\boldsymbol{y}: \ (m, m', \boldsymbol{y}) \in \mathcal{F}(\mathcal{C}_n, Q_{XX'Y})\}. \tag{58}$$

Then, by definition, for $\mathcal{C}_n \in \mathcal{G}_n(Q_{XX'Y})$,

$$\begin{aligned} & \sum_{m,m'} \mathcal{I}\{(\boldsymbol{x}_m \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\} \cdot \frac{|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m \boldsymbol{x}_{m'}) \cap \mathcal{F}(\mathcal{C}_n, Q_{XX'Y}|m, m')|}{|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m \boldsymbol{x}_{m'})|} \\ \geq \ & \exp\{n[2R - I_Q(X;X') - 3\epsilon/2\}, \end{aligned} \tag{59}$$

where we have used the fact that $|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m \boldsymbol{x}_{m'})|$ is the same for all $(\boldsymbol{x}_m \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})$. Putting all this together, we now have:

$$\begin{aligned} & \boldsymbol{E}\left\{[P_{\mathrm{e}}(\mathcal{C}_n)]^{1/\rho}\right\} \\ = \ & \boldsymbol{E}\left[\frac{1}{M}\sum_m \sum_{m' \neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\} + Z_{mm'}(\boldsymbol{y})}\right]^{1/\rho} \\ = \ & \sum_{\mathcal{C}_n} P(\mathcal{C}_n)\left[\frac{1}{M}\sum_m \sum_{m' \neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\} + Z_{mm'}(\boldsymbol{y})}\right]^{1/\rho} \\ \geq \ & \sum_{\mathcal{C}_n \in \mathcal{G}_n} P(\mathcal{C}_n)\left[\frac{1}{M}\sum_{\{Q_{XX'}: \ I_Q(X;X') < 2R - \epsilon\}} \sum_m \sum_{m' \neq m} \mathcal{I}\{(\boldsymbol{x}_m \boldsymbol{x}_{m'}) \in \mathcal{T}(Q_{XX'})\} \times \right. \\ & \sum_{Q_{Y|XX'}} \sum_{\boldsymbol{y} \in \mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m \boldsymbol{x}_{m'}) \cap \mathcal{F}(\mathcal{C}_n, Q_{XX'Y}|m,m')} W(\boldsymbol{y}|\boldsymbol{x}_m) \times \\ & \left. \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\} + Z_{mm'}(\boldsymbol{y})}\right]^{1/\rho} \end{aligned}$$

$$
\begin{aligned}
\geq \;& \sum_{\mathcal{C}_n\in\mathcal{G}_n} P(\mathcal{C}_n)\Bigg[\frac{1}{M}\sum_{\{Q_{XX'}:\ I_Q(X;X')<2R-\epsilon\}}\sum_m\sum_{m'\neq m}\mathcal{I}\{(\boldsymbol{x}_m\boldsymbol{x}_{m'})\in\mathcal{T}(Q_{XX'})\}\times\\
& \sum_{Q_{Y|XX'}}\sum_{\boldsymbol{y}\in\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})\cap\mathcal{F}(\mathcal{C}_n,Q_{XX'Y}|m,m')} W(\boldsymbol{y}|\boldsymbol{x}_m)\times\\
& \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}+\exp\{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]\}}\Bigg]^{1/\rho}\\
\geq \;& \sum_{\mathcal{C}_n\in\mathcal{G}_n} P(\mathcal{C}_n)\Bigg[\frac{1}{M}\sum_{\{Q_{XX'}:\ I_Q(X;X')<2R-\epsilon\}}\sum_m\sum_{m'\neq m}\mathcal{I}\{(\boldsymbol{x}_m\boldsymbol{x}_{m'})\in\mathcal{T}(Q_{XX'})\}\times\\
& \sum_{Q_{Y|XX'}}\frac{|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})\cap\mathcal{F}(\mathcal{C}_n,Q_{XX'Y}|m,m')|}{|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|}\cdot|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|\cdot W(\boldsymbol{y}|\boldsymbol{x}_m)\times\\
& \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}+\exp\{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]\}}\Bigg]^{1/\rho}\\
\geq \;& \sum_{\mathcal{C}_n\in\mathcal{G}_n} P(\mathcal{C}_n)\Bigg[\frac{1}{M}\sum_{\{Q_{XX'Y}:\ I_Q(X;X')<2R-\epsilon\}}\sum_m\sum_{m'\neq m}\mathcal{I}\{(\boldsymbol{x}_m\boldsymbol{x}_{m'})\in\mathcal{T}(Q_{XX'})\}\times\\
& \frac{|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})\cap\mathcal{F}(\mathcal{C}_n,Q_{XX'Y}|m,m')|}{|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|}\cdot|\mathcal{T}(Q_{Y|XX'}|\boldsymbol{x}_m\boldsymbol{x}_{m'})|\cdot W(\boldsymbol{y}|\boldsymbol{x}_m)\times\\
& \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}+\exp\{n[\alpha(R+2\epsilon,Q_Y)+\epsilon]\}}\Bigg]^{1/\rho}\\
\geq \;& P(\mathcal{C}_n)\Bigg[\sum_{\{Q_{XX'Y}:\ I_Q(X;X')<2R-\epsilon\}}\exp\{n[R-I_Q(X;X')-3\epsilon/2]\}\times\\
& \exp\{-n[D(Q_{Y|X}\|W|Q_X)+I_Q(X';Y|X)+\\
& [\max\{g(Q_{XY}),\alpha(R+2\epsilon,Q_Y)+\epsilon\}-g(Q_{X'Y})]_+\}\Bigg]^{1/\rho}\\
\doteq \;& \exp\{-n[E^g_{\rm trc}(R,Q_X)+O(\epsilon)]/\rho\},
\end{aligned}
\tag{60}
$$

which after raising to the power of ρ, gives the desired result, and completes the proof due to the arbitrariness of ϵ.

6 TRC Error Exponents in More General Settings

In this section, we demonstrate that the same analysis technique is applicable to other, more general scenarios of coded communication systems. We briefly outline the analysis and the resulting TRC error exponents in two examples of such scenarios. The first is list decoding where the list size L

is fixed, independently of n. For simplicity, we take $L=2$, but the extension to general L will be straightforward. The second is decoding with an erasure/list option in the framework of Forney [3], where we analyze the exponential rate of the undetected error of the TRC. In both examples, we continue to consider the ensemble of fixed composition codes of type Q_X, and we allow a general decoding metric g, as before.

6.1 List Decoding

Consider a list decoder of list–size L and a deterministic decoder with decoding metric g. Such a decoder outputs the list of the L messages with the highest scores, $g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})$. A list error is the event that the correct codeword is not in the list. For simplicity, we take $L=2$, but the treatment for a general L will be self–evident.

As we observed earlier, in the limit of deterministic decoding ($\beta\to\infty$), for the vast majority of codes, the highest score of an incorrect codeword is typically no larger than $a(R,Q_Y)=\max\{g(Q_{XY}):\ I_Q(X;Y)\le R,\ (Q_Y\odot Q_{X|Y})_X=Q_X\}$ whenever $\boldsymbol{y}\in\mathcal{T}(Q_Y)$. Let us define

$$\Lambda_{\mathrm{L}}(Q_{XX'\tilde{X}})\stackrel{\triangle}{=}\inf_{Q_{Y|XX'\tilde{X}}\in\mathcal{Q}}\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';Y|X)\},\qquad(61)$$

where

$$\mathcal{Q}=\{Q_{Y|XX'\tilde{X}}:\ g(Q_{X'Y})\ge g(Q_{\tilde{X}Y})\ge\max\{g(Q_{XY}),a(R,Q_Y)\}\}.\qquad(62)$$

Then, the TRC list error exponent is given by

$$E_{\mathrm{trcl}}^{g}(R,Q_X)=\inf_{Q_{X'\tilde{X}|X}\in\mathcal{S}(R)}\{\Lambda_{\mathrm{L}}(Q_{XX'\tilde{X}})+I_Q(X;X';\tilde{X})-2R\},\qquad(63)$$

where $I_Q(X;X';\tilde{X})$ is the multi–information, defined as

$$I_Q(X;X';\tilde{X})=H_Q(X)+H_Q(X')+H_Q(\tilde{X})-H_Q(X,X',\tilde{X})\qquad(64)$$

and

$$\begin{aligned}\mathcal{S}(R)\ &=\ \{Q_{X'\tilde{X}|X}:\ \max\{I_Q(X;X'),I_Q(X';\tilde{X}),I_Q(X;\tilde{X})\}<2R,\\ &\qquad I_Q(X;X';\tilde{X})<3R,\ Q_{X'}=Q_{\tilde{X}}=Q_X\}.\end{aligned}\qquad(65)$$

We next provide a brief outline of the derivation, which is largely quite a simple extension of the first part of the proof of Theorem 4, but with a few twists.

$$\begin{aligned}
\boldsymbol{E}\left\{[P_{\mathrm{e}}^{\mathrm{list}}(\mathcal{C}_n)]^{1/\rho}\right\} &\stackrel{\cdot}{\le} \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_m \sum_{m'\ne m}\sum_{\tilde{m}\ne m,m'}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\times\right.\right.\\
&\left.\left.\mathcal{I}\left\{g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\ge g(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\ge \max\{g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}), \max_{\hat{m}\ne m,m'} g(\hat{P}_{\boldsymbol{x}_{\hat{m}}\boldsymbol{y}})\}\right\}\right]^{1/\rho}\right\}\\
&\stackrel{\cdot}{\le} \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_m \sum_{m'\ne m}\sum_{\tilde{m}\ne m,m'}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\times\right.\right.\\
&\left.\left.\mathcal{I}\left\{g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\ge g(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\ge \max\{g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}), a(R-\epsilon,\hat{P}_{\boldsymbol{y}})\}\right\}\right]^{1/\rho}\right\}\\
&\stackrel{\cdot}{\le} \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_{Q_{XX'\tilde{X}}} N(Q_{XX'\tilde{X}})\exp\{-n\Lambda_{\mathrm{L}}(Q_{XX'\tilde{X}})\}\right]^{1/\rho}\right\}\\
&\stackrel{\cdot}{\le} M^{-1/\rho}\sum_{Q_{XX'\tilde{X}}}\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\cdot\exp\{-n\Lambda_{\mathrm{L}}(Q_{XX'\tilde{X}})/\rho\},
\end{aligned}\tag{66}$$

where $N(Q_{XX'\tilde{X}})$ is the number of codeword triplets, $\{\boldsymbol{x}_m, \boldsymbol{x}_{m'}, \boldsymbol{x}_{\tilde{m}}\}$, ($m$, m' and $\tilde{m}$ – all distinct) that fall within $\mathcal{T}(Q_{XX'\tilde{X}})$. Now,

$$\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\} = \sum_{k\ge 1}\Pr\{N(Q_{XX'})=k\}\cdot\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\Big|N(Q_{XX'})=k\}. \tag{67}$$

Consider first the case $I_Q(X;X') > 2R$. Then the sum on the r.h.s. of (67) is dominated by $k \doteq 1$, and so,

$$\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\} \doteq e^{n[2R-I_Q(X;X')]}\cdot\left\{\begin{array}{ll} e^{n[R-I_Q(X,X';\tilde{X})]/\rho} & I_Q(X,X';\tilde{X})<R\\ e^{n[R-I_Q(X,X';\tilde{X})]} & I_Q(X,X';\tilde{X})>R\end{array}\right. \tag{68}$$

which yields

$$\lim_{\rho\to\infty}\left(\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\}\right)^{\rho} \stackrel{\cdot}{\le} \left\{\begin{array}{ll} 0 & I_Q(X,X';\tilde{X})<R\\ 0 & I_Q(X,X';\tilde{X})>R\end{array}\right. = 0. \tag{69}$$

Similarly, $\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\} \doteq 0$ as well when $I_Q(X';\tilde{X}) > 2R$ or $I_Q(X;\tilde{X}) > 2R$. In the case $I_Q(X;X') < 2R$, the sum on the r.h.s. of (67) is dominated by $k \doteq e^{n[2R-I_Q(X;X')]}$, and so,

$$\begin{aligned}
\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\} &\stackrel{\cdot}{\le} e^{n[2R-I_Q(X;X')]s/\rho}\cdot\left(\boldsymbol{E}[N(Q_{\tilde{X}|XX'})]^{1/s}\right)^{s/\rho}\\
&\doteq e^{n[2R-I_Q(X;X')]s/\rho}\cdot\left\{\begin{array}{ll} e^{n[R-I_Q(X,X';\tilde{X})]/\rho} & I_Q(X,X';\tilde{X})<R\\ e^{n[R-I_Q(X,X';\tilde{X})]s/\rho} & I_Q(X,X';\tilde{X})>R\end{array}\right.\\
&= \left\{\begin{array}{ll} e^{n[s(2R-I_Q(X;X'))+R-I_Q(X,X';\tilde{X})]/\rho} & I_Q(X,X';\tilde{X})<R\\ e^{n[3R-I_Q(X;X';\tilde{X})]s/\rho} & I_Q(X,X';\tilde{X})>R\end{array}\right.
\end{aligned}\tag{70}$$

and after optimizing over s,

$$\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\} \quad \stackrel{\cdot}{\le} \quad \begin{cases} e^{n[3R-I_Q(X;X';\tilde{X})]/\rho} & I_Q(X,X';\tilde{X}) < R \\ e^{n[3R-I_Q(X;X';\tilde{X})]/\rho} & I_Q(X,X';\tilde{X}) > R,\ I_Q(X;X';\tilde{X}) < 3R \\ e^{n[3R-I_Q(X;X';\tilde{X})]} & I_Q(X,X';\tilde{X}) > R,\ I_Q(X;X';\tilde{X}) > 3R \end{cases} \tag{71}$$

and so,

$$\begin{aligned} \lim_{\rho\to\infty}\left(\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\}\right)^{\rho} \quad &\stackrel{\cdot}{\le} \quad \begin{cases} e^{n[3R-I_Q(X;X';\tilde{X})]} & I_Q(X,X';\tilde{X}) < R \\ e^{n[3R-I_Q(X;X';\tilde{X})]} & I_Q(X,X';\tilde{X}) > R,\ I_Q(X;X';\tilde{X}) < 3R \\ 0 & R < I_Q(X,X';\tilde{X}),\ I_Q(X;X';\tilde{X}) > 3R \end{cases} \\ &= \quad \begin{cases} e^{n[3R-I_Q(X;X';\tilde{X})]} & I_Q(X;X';\tilde{X}) < 3R \\ 0 & I_Q(X;X';\tilde{X}) > 3R \end{cases} \end{aligned} \tag{72}$$

A similar derivation applies to the two other combinations of two out the three random variables, X, X' and $\tilde{X}$. In summary,

$$\lim_{\rho\to\infty}\left(\boldsymbol{E}\{[N(Q_{XX'\tilde{X}})]^{1/\rho}\}\right)^{\rho} \stackrel{\cdot}{\le} \begin{cases} e^{n[3R-I_Q(X;X';\tilde{X})]} & Q_{X'\tilde{X}|X} \in \mathcal{S}(R) \\ 0 & \text{elsewhere} \end{cases} \tag{73}$$

and the desired result follows similarly as before.

6.2 Decoding with an Erasure/List Option

Consider the following generalized version of Forney's erasure/list decoder [3], which for a given parameter T, decides in favor of message m whenever

$$\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}}{\sum_{m'\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}} \ge e^{nT}, \tag{74}$$

and erases if no message m satisfies this inequality.

We now define

$$\begin{aligned} &\Lambda(Q_{XX'},R,T) \\ =\quad &\min_{\{Q_{Y|XX'}:\ g(Q_{X'Y})-\max\{g(Q_{XY}),\alpha(R,Q_Y)\}\ge T\}}\left\{D(Q_{Y|X}\|W|Q_X)+I_Q(X';Y|X)\right\}. \end{aligned} \tag{75}$$

and then we argue that TRC undetected error exponent is given by

$$E^{g}_{\text{trc-ue}}(R,T,Q_X) = \min_{\{Q_{XX'}:\ I_Q(X;X')\le 2R,\ Q_{X'}=Q_X\}}[\Lambda(Q_{XX'},R,T)+I_Q(X;X')-R]. \tag{76}$$

The outline is for the derivation is as follows. For a given code $\mathcal{C}_n$, the probability of undetected error is given by

$$\begin{aligned}
P_{\text{ue}}(\mathcal{C}_n) &= \frac{1}{M}\sum_{m=0}^{M-1}\Pr\bigcup_{m'\neq m}\left\{\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\sum_{\tilde{m}\neq m'}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}\geq e^{nT}\right\}\\
&= \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\Pr\left\{\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\sum_{\tilde{m}\neq m'}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}\geq e^{nT}\right\}\\
&= \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\Pr\left\{\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}})\}+\sum_{\tilde{m}\neq m,m'}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}\geq e^{nT}\right\}\\
&\stackrel{\triangle}{=} \frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x}_m)\cdot\mathcal{I}\left\{\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}})\}+Z_{mm'}(\boldsymbol{y})}\geq e^{nT}\right\}. \qquad (77)
\end{aligned}$$

Thus, using the same considerations as before,

$$\begin{aligned}
&\boldsymbol{E}\left\{[P_{\text{ue}}(\mathcal{C}_n)]^{1/\rho}\right\}\\
&= \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x}_m)\cdot\mathcal{I}\left\{\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}})\}+Z_{mm'}(\boldsymbol{y})}\geq e^{nT}\right\}\right]^{1/\rho}\right\}\\
&\stackrel{\cdot}{\leq} \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x}_m)\times\right.\right.\\
&\quad\left.\left.\mathcal{I}\left\{\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}})\}+\exp\{n\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})\}}\geq e^{nT}\right\}\right]^{1/\rho}\right\}\\
&\stackrel{\cdot}{=} \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\sum_{\boldsymbol{y}}W(\boldsymbol{y}|\boldsymbol{x}_m)\times\right.\right.\\
&\quad\left.\left.\mathcal{I}\left\{g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})-\max\{g(\hat{P}_{\boldsymbol{x}_{m}\boldsymbol{y}}),\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})\geq T\right\}\right]^{1/\rho}\right\}. \qquad (78)
\end{aligned}$$

By the method of types [2], the inner–most sum is of the exponential order of $\exp\{-n\Lambda(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R,T)\}$, and so,

$$\begin{aligned}
\boldsymbol{E}\left\{[P_{\text{ue}}(\mathcal{C}_n)]^{1/\rho}\right\} &\stackrel{\cdot}{\leq} \boldsymbol{E}\left\{\left[\frac{1}{M}\sum_{m=0}^{M-1}\sum_{m'\neq m}\exp\{-n\Lambda(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}},R,T)\}\right]^{1/\rho}\right\}\\
&= M^{-1/\rho}\boldsymbol{E}\left\{\left[\sum_{Q_{XX'}}N(Q_{XX'})\exp\{-n\Lambda(Q_{XX'},R,T)\}\right]^{1/\rho}\right\}\\
&\stackrel{\cdot}{\leq} M^{-1/\rho}\sum_{Q_{XX'}}\boldsymbol{E}\left\{[N(Q_{XX'})]^{1/\rho}\right\}\cdot\exp\{-n\Lambda(Q_{XX'},R,T)/\rho\}, \qquad (79)
\end{aligned}$$

which yields, similarly as before,

$$\begin{aligned}
&\lim_{\rho\to\infty}\left(\boldsymbol{E}\left\{[P_{\mathrm{ue}}(\mathcal{C}_n)]^{1/\rho}\right\}\right)^\rho\\
\doteq\;&\exp\left\{-n\min_{\{Q_{XX'}:\;I_Q(X;X')\le 2R,\;Q_{X'}=Q_X\}}[\Lambda(Q_{XX'},R,T)+I_Q(X;X')-R]\right\}\\
=\;&\exp\{-nE^g_{\text{trc-ue}}(R,T,Q_X)\}. \qquad (80)
\end{aligned}$$

Acknowledgement

Interesting discussions with Anelia Somekh–Baruch, in the early stages of this work, are acknowledged with thanks.

References

[1] A. Barg and G. D. Forney, Jr., "Random codes: minimum distances and error exponents," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2568–2573, September 2002.

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

[3] G. D. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inform. Theory*, vol. IT–14, no. 2, pp. 206–220, March 1968.

[4] G. D. Forney, Jr., "On exponential error bounds for random codes on the BSC," course notes, M.I.T., Spring 2005. Available on–line at:
`http://web.mit.edu/6.441/spring05/reading/Forney_ExpEBBSC.pdf`

[5] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.

[6] Y. Kabashima, "How could the replica method improve accuracy of performance assessment of channel coding?" *Proc. Int. Workshop on Statistical–Mechanical Informatics*, Sept. 14–17, 2008, Sendai, Japan. `arXiv:0808.0548v1 [cs.IT] 5 Aug 2008`

[7] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, nos. 1–2, pp. 1–212, 2009.

[8] N. Merhav, "The generalized stochastic likelihood decoder: random coding and expurgated bounds," *IEEE Trans. Inform. Theory*, vol. 63, no. 8, pp. 5039–5051, August 2017. See also correction at `https://arxiv.org/pdf/1707.03987.pdf`

[9] N. Merhav, "Ensemble performance of biometric authentication systems based on secret key generation," submitted for publication, July 2017.
`https://arxiv.org/pdf/1707.09108.pdf`

[10] T. Mora and O. Rivoire, "Statistical mechanics of error exponents for error–correcting codes," *Phys. Rev. E*, vol. 74 056110, 2006.

[11] J. Scarlett, A. Martinéz and A. G. i Fábregas, "The likelihood decoder: error exponents and mismatch," *Proc. 2015 IEEE International Symposium on Information Theory (ISIT 2015)*, pp. 86–90, Hong Kong, June 2015.

[12] N. S. Skantzos, J. van Mourik, Y. Kabashima, and D. Saad, "Average error exponent in Gallager low–density parity–check codes," *J. Phys. A: Math. Gen.*, vol. 36, pp. 11131–11141, 2003.

[13] E. C. Song, P. Cuff and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inform. Theory*, vol. 62, no. 4, pp. 1836–1849, April 2016.

[14] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw–Hill, New York 1979.

[15] M. H. Yassaee, M. R. Aref and A. Gohari, "A technique for deriving one–shot achievability results in network information theory," *Proc. 2013 IEEE International Symposium on Information Theory (ISIT 2013)*, pp. 1287–1291, July 2013. Also, available on–line at `http://arxiv.org/abs/1303.0696`.