IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

The Generalized Stochastic Likelihood Decoder: Random Coding and Expurgated Bounds

Neri Merhav

CCIT Report #891
December 2015

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #891 December 2015

The Generalized Stochastic Likelihood Decoder: Random Coding and Expurgated Bounds

Neri Merhav *

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

The likelihood decoder is a stochastic decoder that selects the decoded message at random, using the posterior distribution of the true underlying message given the channel output. In this work, we study a generalized version of this decoder where the posterior is proportional to a general function that depends only on the joint empirical distribution of the output vector and the codeword. This framework allows both mismatched versions and universal (MMI) versions of the likelihood decoder, as well as the corresponding ordinary deterministic decoders, among many others. We provide a direct analysis method that yields the exact random coding exponent (as opposed to separate upper bounds and lower bounds that turn out to be compatible, which were derived earlier by Scarlett *et al.*). We also extend the result from pure channel coding to combined source and channel coding (random binning followed by random channel coding) with side information available to the decoder. Finally, returning to pure channel coding, we derive also an expurgated exponent for the stochastic likelihood decoder, which turns out to be at least as tight (and in some cases, strictly so) as the classical expurgated exponent of the maximum likelihood decoder, even though the stochastic likelihood decoder is suboptimal.

Index Terms Stochastic decoder, likelihood decoder, random coding exponent, expurgated exponent, random binning, source–channel coding.

*This research is partially supported by the Israel Science Foundation (ISF), grant no. 412/12.

1 Introduction

The likelihood decoder for channel coding is a stochastic decoder that selects the decoded message at random under the posterior distribution of the correct message given the received channel output vector. The likelihood decoder has recently received some attention, with the primary motivation that it lends itself to considerably simpler derivations of asymptotic upper bounds on the error probability in a variety of problems of network information theory [17]. Owing to the duality between source encoding and channel decoding, the likelihood encoder was also studied in the context of rate–distortion coding [16].

More recently, in [15] exact error exponents have been derived for a mismatched version of the likelihood decoder, assuming a discrete memoryless channel (DMC) and using the ensembles i.i.d. and constant composition codes. It was shown in [15], among many other results, that in the special case of the (matched) likelihood decoder, the random coding error exponents achieved, in both ensembles, are exactly the same as the corresponding random coding error exponents of the optimal maximum likelihood (ML) decoder.

The focus of this work is on further developments concerning the exact error exponent analysis of [15], as well as on extensions and refinements of this analysis in several directions. In particular, the main contributions of this work are the following.

1. Allowing a more general family of stochastic likelihood decoders, according to which the probability of deciding on a given message is proportional to a general exponential function of the joint empirical distribution of the codeword and the received channel output vector. This is more general than the mismatched likelihood decoder of [15].

2. Providing a direct, exponentially tight derivation of the random coding exponent in a single analysis, instead of the separate upper and lower bounds of [15] (which turn out to coincide). Hence we believe that this analysis is somewhat simpler, at least conceptually.

3. Extending the scope to a situation of source–channel coding with side information at the decoder, where the source coding part is based on random binning (similarly as in [11]), thus covering a variety of settings of theoretical and practical interest, including joint source–channel coding with side information.

4. Returning to pure channel coding, we derive also an expurgated bound. We point out that when this result is applied to the ordinary likelihood decoder (which uses the real posterior probability of each message), the resulting expurgated bound is guaranteed to be *at least as tight* as the classical expurgated bound due to Csiszár, Körner and Marton [2], [3], which in turn is at least as tight as Gallager's expurgated bound [5]. This is in spite of the fact that the likelihood decoder analyzed is suboptimal. We also demonstrate that the new expurgated bound may strictly improve on the classical expurgated bound at least at high rates.

Finally, a few comments are in order regarding the error exponent analysis. The analysis technique used is primarily the type class enumeration method [7, Chap. 6], which has already proved quite useful as a tool for obtaining exponentially tight random coding bounds in various contexts (see, e.g., [8], [9], [10], for a sample). When it comes to the extension of the setup to source–channel coding with side information, the ensemble of codes in our setting combines random binning (for

the source coding part) and random coding (for the channel coding part), which is somewhat more involved than ordinary error exponent analyses that involve either one but not both. This requires quite a careful analysis, which similarly as in [11], is carried out in two steps: first, we take the average probability of error over the ensemble of random binning codes, for a given channel code, and at the second step, we average over the ensemble of channel codes.

The remaining part of the paper is organized as follows. In Section 2, we establish notation conventions, provide some background, and define the objectives of this paper more accurately. In Section 3, we re-derive the exact random coding exponent of [15] in an alternative way, as described above. Section 4 is devoted to the extension to source–channel coding with side information, and finally, Section 5 is about the expurgated bound.

2 Notation Conventions, Background and Objectives

2.1 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, $Q_{Y|X}$, and so on. For example, the joint distribution of (X, Y), induced by Q_X and $Q_{Y|X}$, will be denoted by Q_{XY} and the corresponding marginal of Y will be denoted by Q_Y. When there is no room for ambiguity, the subscripts will be omitted. When we wish to refer to the joint distribution induced by the input assignment Q_X and a conditional distribution other than $Q_{Y|X}$, say $W_{Y|X}$, we denote it by $Q_X \times W$, or simply $Q \times W$. In this case, the marginal of Y, that is induced by $Q \times W$, will be denoted by $(Q \times W)_Y$. The probability of an event $\mathcal{E}$ will be denoted by $\Pr\{\mathcal{E}\}$, and the expectation operator with respect to (w.r.t.) a probability distribution $Q \times P$ will be denoted by $\boldsymbol{E}_Q\{\cdot\}$. Again, the subscript will be omitted if the underlying probability distribution is clear from the context. Concerning the notation of information measures, the entropy of a random variable X, with a distribution Q, will be denoted by $H_Q(X)$. Similarly, for a joint distribution Q of (X, Y), the conditional entropy will be denoted by $H_Q(X|Y)$, the mutual information will be denoted by $I_Q(X;Y)$, and so on. When we wish to focus our emphasis on the dependence of the mutual information only upon the underlying joint distribution, we denote it instead by $I(Q)$. The relative entropy (or the Kullback–Leibler divergence) between two conditional distributions, $Q_{Y|X}$ and $W_{Y|X}$ (or simply W), weighted by the input assignment Q_X, will be denoted and defined by

$$D(Q_{Y|X}\|W|Q_X) = D(Q_{XY}\|Q_X \times W) = \sum_{x\in\mathcal{X}} Q(x) \sum_{y\in\mathcal{Y}} Q(y|x) \log \frac{Q(y|x)}{W(y|x)}, \tag{1}$$

where here and throughout the sequel, logarithms will be understood to be defined with respect to the natural basis.

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, $a_n \dot{\leq} b_n$ means that $\limsup_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} \leq 0$, and so on. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathcal{I}\{E\}$. The notation $[x]_+$ will stand for $\max\{0, x\}$.

The empirical distribution of a sequence $\boldsymbol{x} \in \mathcal{X}^n$, which will be denoted by $\hat{P}_{\boldsymbol{x}}$, is the vector of relative frequencies $\hat{P}_{\boldsymbol{x}}(x)$ of each symbol $x \in \mathcal{X}$ in $\boldsymbol{x}$. The type class of $\boldsymbol{x} \in \mathcal{X}^n$, denoted $\mathcal{T}(\boldsymbol{x})$, is the set of all vectors $\boldsymbol{x}'$ with $\hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}$. When we wish to emphasize the dependence of the type class on the empirical distribution $\hat{P}$, we will denote it by $\mathcal{T}(\hat{P})$. 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 $\hat{H}_{\boldsymbol{x}}(X)$. An alternative notation, following the conventions described in the previous paragraph, is $H(\hat{P}_{\boldsymbol{x}})$. 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, $\hat{P}_{\boldsymbol{xy}}$ would be the joint empirical distribution of $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}_{i=1}^n$, $\mathcal{T}(\boldsymbol{x}, \boldsymbol{y})$ or $\mathcal{T}(\hat{P}_{\boldsymbol{xy}})$, will denote the joint type class of $(\boldsymbol{x}, \boldsymbol{y})$, $\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})$ or $\mathcal{T}(\hat{P}_{\boldsymbol{x}|\boldsymbol{y}}|\boldsymbol{y})$, will stand for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X, Y)$ or $H(\hat{P}_{\boldsymbol{xy}})$, will designate the empirical joint entropy of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X|Y)$ will be the empirical conditional entropy, and $\hat{I}_{\boldsymbol{xy}}(X;Y)$ (or alternatively, $I(\hat{P}_{\boldsymbol{xy}})$) will denote empirical mutual information, etc.

2.2 Background – The Generalized Likelihood Decoder

Consider a DMC, designated by a matrix of single–letter input–output transition probabilities $\{W(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$. Here the channel input symbol x takes on values in a finite input alphabet $\mathcal{X}$, and the channel output symbol y takes on values in a finite output alphabet $\mathcal{Y}$. When the channel is fed by a vector $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, it outputs a vector $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ according to

$$W(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^{n} W(y_t|x_t). \tag{2}$$

A code $\mathcal{C}_n \subseteq \mathcal{X}^n$ is a collection of $M = e^{nR}$ channel input vectors, $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$, R being the coding rate in nats per channel use. It is assumed that all messages, $m = 0, 1, \ldots .M-1$, are equally likely.

As is very common in the information theory literature, we will consider, throughout most of this work, the random coding regime. The random coding ensemble considered (here, as well as as in [15]) is the ensemble of constant composition codes, where each codeword is drawn independently under the uniform distribution within a given type class $\mathcal{T}(Q_X)$. Once the code has been randomly selected, it is revealed to both the encoder and the decoder.

When the transmitter wishes to convey a message m, it transmits the corresponding code-vector $\boldsymbol{x}_m$ via the channel, which in turn, stochastically maps it into an n–vector $\boldsymbol{y}$ according to (2). Upon receiving $\boldsymbol{y}$, the stochastic *likelihood decoder* randomly selects the estimated message $\hat{m}$ according to the induced posterior distribution of the transmitted message, i.e.,

$$\Pr\{\hat{m} = m_0|\boldsymbol{y}\} = \Pr\{m = m_0|\boldsymbol{y}\} = \frac{W(\boldsymbol{y}|\boldsymbol{x}_{m_0})}{\sum_{m=0}^{M-1} W(\boldsymbol{y}|\boldsymbol{x}_m)}. \tag{3}$$

Inspired by earlier work on mismatched decoding (see, e.g., [4], [6], [13]), Scarlett *et al.* [15] studied a mismatched version of the likelihood decoder, which is defined similarly as in (3), but with a mismatched DMC W' replacing the true one, W. The main results of [15] are single–letter formulae for the exact random coding error exponent of the mismatched likelihood decoder. Specifically, the random coding exponent derived in [15] (see Lemma 1 therein) is given by

$$\begin{aligned} E(R) &= \min_{Q_{Y|X}} \min_{\{Q'_{Y|X}:\ (Q_X \times Q'_{Y|X})_Y = Q_Y\}} \Big\{ D(Q_{Y|X} \| W | Q_X) + \\ &\qquad [I(Q_X \times Q'_{Y|X}) + [\boldsymbol{E}_Q \log W'(Y|X) - \boldsymbol{E}_{Q_X \times Q'_{Y|X}} \log W'(Y|X)]_+ - R]_+ \Big\}. \end{aligned} \tag{4}$$

One of the interesting conclusions in [15] is that in the special case of the regular matched likelihood decoder ($W' = W$), this expression of the random coding error exponent coincides with that of the classical ML decoder.

2.3 Objectives and Main Contributions

The generalized likelihood decoder (GLD) to be considered in this work, is defined according to

$$\Pr\{\hat{m} = m_0 | \boldsymbol{y}\} = \frac{\exp\{n g(\hat{P}_{\boldsymbol{x}_{m_0}\boldsymbol{y}})\}}{\sum_{m=0}^{M-1} \exp\{n g(\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}})\}}, \tag{5}$$

where $\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}$ is the empirical distribution of $(\boldsymbol{x}_m, \boldsymbol{y})$ (whose X-marginal, $\hat{P}_{\boldsymbol{x}}$, coincides with Q) and g is a given continuous, real valued functional of this empirical distribution.

This generalized likelihood decoder covers several important special cases. Obviously, the choice

$$g(\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}) = \sum_{x,y} \hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}(x,y) \log W(y|x) \tag{6}$$

corresponds to the ordinary likelihood decoder. Slightly more generally, one may introduce a parameter $\beta \geq 0$ and define

$$g(\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}) = \beta \sum_{x,y} \hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}(x,y) \log W(y|x). \tag{7}$$

Here, β controls the degree of skewedness of the distribution (5), in the spirit of the notion of finite–temperature decoding [14]: while $\beta = 1$ corresponds to the usual stochastic likelihood decoder, $\beta \to \infty$ leads to the traditional (deterministic) ML decoder. Likewise,

$$g(\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}) = \beta \sum_{x,y} \hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}(x,y) \log W'(y|x) \tag{8}$$

defines a family of mismatched likelihood decoders, bridging between the mismatched likelihood decoder of [15] and the ordinary, deterministic mismatched decoder (although the parameter β might as well be absorbed in W' in the form of a power of W'). Yet another important example is

$$g(\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}) = \beta I(\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}), \tag{9}$$

which is a parametric family of stochastic maximum mutual information (MMI) decoders, where once again, $\beta \to \infty$ yields the ordinary MMI universal decoder [2].

The main contributions in this paper are the following.

1. Allowing the above described more general family of stochastic likelihood decoders (5) with a general function g. While technically, this extension is quite straightforward,[1] it is important to allow g to be a general (not necessarily linear) functional of the joint empirical distribution, as it covers, for example, the important class of MMI likelihood decoders with g defined as in (9).

2. While in [15] eq. (4) is derived by separate analyses of an upper bound and a matching lower bound, here we provide directly an exponentially tight derivation of the random coding exponent in a single analysis. We believe that this analysis is somewhat simpler, at least conceptually.

3. Extending the scope to a situation of source–channel coding with side information at the decoder, where the source coding part is based on random binning (similarly as in [11]), thus covering a variety of settings of theoretical and practical interest, including pure source coding, pure channel coding, joint/separate source–channel coding with and without side information, systematic coding, etc. Here, the distribution of the decoded source message given the channel output $\boldsymbol{y}$ is assumed to be proportional to the product of two functions, the first depending on the joint type of the source vector $\boldsymbol{u}$ and the side information $\boldsymbol{v}$, and the second one depends on the corresponding code word $\boldsymbol{x}(\boldsymbol{u})$ and the channel output $\boldsymbol{y}$.

4. Returning to pure channel coding, we derive also an expurgated error exponent. An interesting point to consider is that when this is applied to the ordinary likelihood decoder (3), the resulting expurgated bound is guaranteed to be *at least as tight* as the classical expurgated bound due to Csiszár, Körner and Marton [2], [3], and this is in spite of the fact that the likelihood decoder analyzed is suboptimal. In this context, we study the example of the z–channel and demonstrate that the new expurgated bound strictly improves on the classical expurgated bound at high rates.

3 Another Derivation of the Random Coding Exponent

In this section, we provide an alternative derivation of $E(R)$, given in (4), which is different from the one in [15], as described in item no. 1 above.

Assuming, without loss of generality, that message $m = 0$ was transmitted, the average probability of error of the GLD is given by

$$\begin{aligned}\bar{P}_{\mathrm{e}} &= \boldsymbol{E}\left\{\frac{\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{Y}})\}}{\sum_{m=0}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{Y}})\}}\right\} \\ &= \boldsymbol{E}\left[\boldsymbol{E}\left\{\frac{\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{Y}})\}}{\sum_{m=0}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{Y}})\}}\Big|\boldsymbol{X}_0,\boldsymbol{Y}\right\}\right], \qquad (10)\end{aligned}$$

where the inner expectation is taken w.r.t. the randomness of the incorrect codewords, $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_{M-1}$, and the outer expectation is taken w.r.t. the randomness of the transmitted codeword $\boldsymbol{X}_0$ and the channel output $\boldsymbol{Y}$. We first address the inner expectation for given realizations $(\boldsymbol{X}_0,\boldsymbol{Y}) = (\boldsymbol{x}_0,\boldsymbol{y})$.

[1] Just replace $\boldsymbol{E}_Q \log W'(Y|X)$ and $\boldsymbol{E}_{Q_X\times Q'_{Y|X}} \log W'(Y|X)$, of (4), by $g(Q)$ and $g(Q_X\times Q'_{Y|X})$, respectively,

Let $N_{\boldsymbol{y}}(Q')$ denote the number of codewords, other than $\boldsymbol{x}_0$, whose joint empirical distribution with $\boldsymbol{y}$ is given by Q'. Then,

$$\begin{aligned}
\bar{P}_{\mathrm{e}}(\boldsymbol{x}_0,\boldsymbol{y}) &= \boldsymbol{E}\left\{\frac{\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})\}+\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}}\right\}\\
&= \int_0^1 \Pr\left\{\frac{\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})\}+\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}}\geq t\right\}\mathrm{d}t\\
&= n\cdot\int_0^\infty e^{-n\theta}\Pr\left\{\frac{\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})\}+\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}}\geq e^{-n\theta}\right\}\mathrm{d}\theta\\
&= n\cdot\int_0^\infty e^{-n\theta}\Pr\left\{(1-e^{-n\theta})\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}\geq e^{-n\theta}\exp\{ng(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})\}\right\}\mathrm{d}\theta\\
&\doteq \int_0^\infty e^{-n\theta}\Pr\left\{\sum_{m=1}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{X}_m\boldsymbol{y}})\}\geq\exp\{n[g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})-\theta]\}\right\}\mathrm{d}\theta\\
&\doteq \int_0^\infty e^{-n\theta}\Pr\left\{\sum_{Q'}N_{\boldsymbol{y}}(Q')e^{ng(Q')}\geq\exp\{n[g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})-\theta]\}\right\}\mathrm{d}\theta\\
&\doteq \int_0^\infty e^{-n\theta}\Pr\left\{\max_{Q'}N_{\boldsymbol{y}}(Q')e^{ng(Q')}\geq\exp\{n[g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})-\theta]\}\right\}\mathrm{d}\theta\\
&\doteq \int_0^\infty e^{-n\theta}\Pr\bigcup_{Q'}\left\{N_{\boldsymbol{y}}(Q')e^{ng(Q')}\geq\exp\{n[g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})-\theta]\}\right\}\mathrm{d}\theta\\
&\doteq \sum_{Q'}\int_0^\infty e^{-n\theta}\Pr\left\{N_{\boldsymbol{y}}(Q')e^{ng(Q')}\geq\exp\{n[g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})-\theta]\}\right\}\mathrm{d}\theta\\
&\doteq \max_{Q'}\int_0^\infty e^{-n\theta}\Pr\left\{N_{\boldsymbol{y}}(Q')e^{ng(Q')}\geq\exp\{n[g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})-\theta]\}\right\}\mathrm{d}\theta\\
&\doteq \max_{Q'}\int_0^\infty e^{-n\theta}\Pr\left\{N_{\boldsymbol{y}}(Q')\geq\exp\{n[g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}})-g(Q')-\theta]\}\right\}\mathrm{d}\theta\\
&\stackrel{\Delta}{=} \max_{Q'}\bar{P}_{\mathrm{e}}(\boldsymbol{x}_0,\boldsymbol{y},Q'),
\end{aligned}\tag{11}$$

where the unions, summations and the maximizations over $\{Q'\}$ are understood to be taken over all possible empirical distributions of sequence pairs of length n, whose X-marginals coincide with Q_X. Henceforth, for the sake of simplicity and consistency with the earlier defined notation, we replace the notation $\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}}$ by Q. Now, given $\boldsymbol{y}$, $N_{\boldsymbol{y}}(Q')$ is a binomial random variable with e^{nR} trials and success rate of the exponential order of $e^{-nI(Q')}$. Therefore, using the techniques of [7, Section 6.3]

$$\Pr\left\{N_{\boldsymbol{y}}(Q')\geq\exp\{n[g(Q)-g(Q')-\theta]\}\right\}\doteq e^{-nE_1(\theta,Q,Q',R)}\tag{12}$$

where

$$\begin{aligned}
E_1(\theta,Q,Q',R) &= \begin{cases}[I(Q')-R]_+ & g(Q)-g(Q')-\theta\leq[R-I(Q')]_+\\ \infty & \text{elsewhere}\end{cases}\\
&= \begin{cases}[I(Q')-R]_+ & \theta\geq g(Q)-g(Q')-[R-I(Q')]_+\\ \infty & \text{elsewhere}\end{cases}
\end{aligned}\tag{13}$$

and so,

$$\begin{aligned}\bar{P}_{\rm e}(\boldsymbol{x}_0,\boldsymbol{y},Q) &= \int_0^\infty e^{-n\theta}\Pr\left\{N_{\boldsymbol{y}}(Q')\geq\exp\{n[g(Q)-g(Q')-\theta]\}\right\}{\rm d}\theta\\ &\dot{=} \int_{[g(Q)-g(Q')-[R-I(Q')]_+]_+}^\infty e^{-n\theta}\cdot e^{-n[I(Q')-R]_+}{\rm d}\theta\\ &\dot{=} \exp\left\{-n([I(Q')-R]_+ + [g(Q)-g(Q')-[R-I(Q')]_+]_+)\right\}\\ &\stackrel{\Delta}{=} e^{-nE_2(Q,Q',R)}\end{aligned} \tag{14}$$

where $E_2(Q,Q',R)$ can also be written as

$$E_2(Q,Q',R)=\left\{\begin{array}{ll}[I(Q')-R+g(Q)-g(Q')]_+ & R\geq I(Q')\\ I(Q')-R+[g(Q)-g(Q')]_+ & R<I(Q')\end{array}\right. \tag{15}$$

As explained briefly in [15], this expression can be simplified as follows. First, for a given $a\in\mathbb{R}$ and $b\in\mathbb{R}^+$, consider the identity[2] $[a-b]_+=[[a]_+-b]_+$, and applying it to the first line of (15) with $a=g(Q)-g(Q')$ and $b=R-I(Q')$. Then, the first line of (15) can also be expressed as $[I(Q')-R+[g(Q)-g(Q')]_+]_+$. Now, since the second line of (15) is non–negative, it can also be expressed as $[I(Q')-R+[g(Q)-g(Q')]_+]_+$. Therefore

$$E_2(Q,Q',R)=[I(Q')-R+[g(Q)-g(Q')]_+]_+ \tag{16}$$

regardless of whether $R\geq I(Q')$ or $R<I(Q')$. Next, define

$$E_3(Q,R)=\min_{Q'}E_2(Q,Q',R), \tag{17}$$

where the minimization is over all joint distributions $\{Q'\}$ whose X–marginal is consistent with Q_X and whose Y–marginal agrees with Q_Y. Finally, the error exponent of the GLD is given by

$$E(R)=\min_Q[D(Q\|Q_X\times W)+E_3(Q,R)], \tag{18}$$

where the minimization is over all joint distributions $\{Q\}$ whose X-marginal is Q_X. This recovers the expression (4) derived in [15].

Several comments are now in order.

1. First, observe that for $g(Q)=I(Q)$, we have

$$E_2(Q,Q',R)=[I(Q')-R+[I(Q)-I(Q')]_+]_+=[\max\{I(Q),I(Q')\}-R]_+\geq[I(Q)-R]_+ \tag{19}$$

yielding

$$E(R)\geq\min_Q\{D(Q\|Q_X\times W)+[I(Q)-R]_+\}, \tag{20}$$

which is exactly the random coding error exponent of the ML decoder [2]. This holds true also for $g(Q)=\beta I(Q)$, provided that $\beta\geq 1$, since the exponent is monotonically increasing in β, but on the other hand, cannot exceed the exponent of the ML decoder. This is in analogy to the case

[2] To see why this identity is true, observe that if $a>b$, then $a>0$, which means $a=[a]_+$ and the identity obviously holds. Otherwise, if $a\leq b$, then $[a]_+\leq b$ as well (again, due to the positivity of b), in which case both $a-b$ and $[a]_+-b$ are non–positive, and so $[a-b]_+=[[a]_+-b]_+=0$.

$g(Q) = \beta \sum_{x,y} Q(x,y) \ln W(y|x)$, which was shown in [15] to achieve the same exponent as the ML decoder even for $\beta = 1$, and therefore also for every $\beta \geq 1$.

2. The highest achievable rate is calculated as follows: we seek a condition on R such that $E(R) > 0$, namely, for all Q and all Q' (consistent with Q),

$$D(Q\|Q_X \times W) + [I(Q') - R + [g(Q) - g(Q')]_+]_+ > 0 \tag{21}$$

or, equivalently,

$$\max_{s,t\in[0,1]} \big\{D(Q\|Q_X \times W) + s[I(Q') - R + t[g(Q) - g(Q')]]\big\} > 0. \tag{22}$$

In other words, we need that for every Q and Q', there exists s and t, both in $[0,1]$, such that

$$D(Q\|Q_X \times W) + s[I(Q') - R + t[g(Q) - g(Q')]] > 0. \tag{23}$$

i.e.,

$$\forall Q, Q' \ \exists \ s,t \in [0,1]^2 : \ R < I(Q') + t[g(Q) - g(Q')] + \frac{D(Q\|Q_X \times W)}{s} \tag{24}$$

which means

$$\begin{aligned} R &< R_0 \stackrel{\Delta}{=} \min_{Q} \min_{Q'} \max_{s,t\in[0,1]} \left\{ I(Q') + t[g(Q) - g(Q')] + \frac{D(Q\|Q_X \times W)}{s} \right\} \\ &= \min_{Q'} \min_{Q} \begin{cases} I(Q') + [g(Q_X \times W) - g(Q')]_+ & Q_{Y|X} = W \\ \infty & Q_{Y|X} \neq W \end{cases} \\ &= \min_{Q'} \{I(Q') + t[g(Q_X \times W) - g(Q')]_+\}. \end{aligned} \tag{25}$$

where it should be kept in mind that the minimization is over all $\{Q'\}$ whose X–marginal is Q_X and whose Y-marginal is consistent with $Q_X \times W$. Obviously,

$$R_0 \leq \min_{Q} \{I(Q) + t[g(Q_X \times W) - g(Q)]_+\} = \min\{I(Q) : \ g(Q) \leq g(Q_X \times W)\} \leq I(Q_X \times W). \tag{26}$$

A lower on the achievable rate bound can be obtained by

$$R_0 \geq \max_{t\in[0,1]} \min_{Q} \{I(Q) - tg(Q) + tg(Q_X \times W)\}, \tag{27}$$

which is tight when $I(Q) - tg(Q)$ is convex in Q for every $t \in [0,1]$, as is the case when g is linear in Q and when $g(Q) = I(Q)$.

4 Extension to Source–Channel Coding With Side Information

Consider the communication system depicted in Fig. 1. Let $(\boldsymbol{U}, \boldsymbol{V}) = \{(U_t, V_t)\}_{t=1}^n$ be n independent copies of a pair of random variables, $(U,V) \sim P_{UV}$, taking on values in finite alphabets, $\mathcal{U}$ and $\mathcal{V}$, respectively. The vector $\boldsymbol{U}$ will designate the source vector to be encoded, whereas the vector $\boldsymbol{V}$ will serve as correlated side information, available to the decoder. When a given realization $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathcal{U}^n$, of the finite alphabet source vector $\boldsymbol{U}$, is fed into the system, it is encoded into one out of $M = e^{nR}$ bins, selected independently at random for every member of $\mathcal{U}^n$. Here, $R > 0$ is referred to as the *binning rate*. The bin index $j = b(\boldsymbol{u})$ is mapped into a channel input

vector $\boldsymbol{x}(j) \in \mathcal{X}^n$, which in turn is transmitted across the channel W. The decoder estimates $\boldsymbol{u}$ based on the channel output $\boldsymbol{y}$ and the side information sequence $\boldsymbol{v}$, which is a realization of $\boldsymbol{V}$. As before, the various codewords $\{\boldsymbol{x}(j)\}_{j=1}^{M}$ are selected independently at random under the uniform distribution across a given type class $\mathcal{T}(Q_X)$. With a slight abuse of notation, we will sometimes denote $\boldsymbol{x}(j) = \boldsymbol{x}[f(\boldsymbol{u})]$ by $\boldsymbol{x}[\boldsymbol{u}]$.

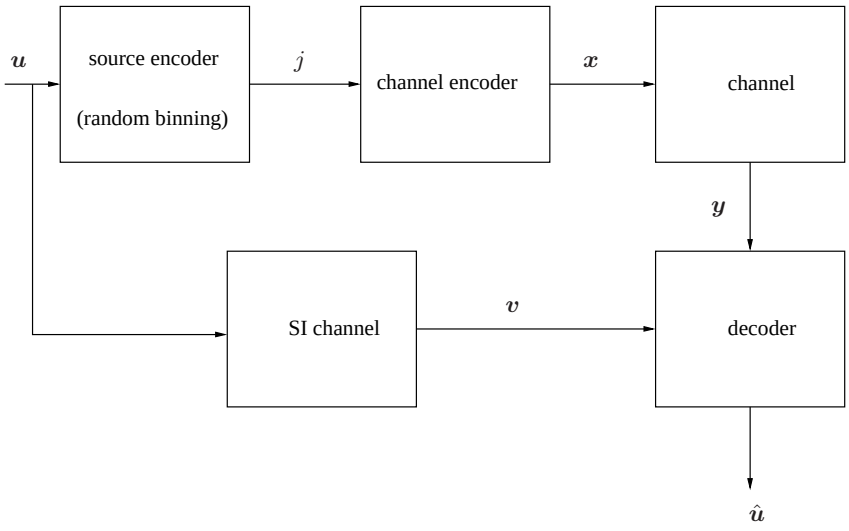

Figure 1: Slepian–Wolf source coding, followed by channel coding. The source $\boldsymbol{u}$ is source–channel encoded, whereas the correlated SI $\boldsymbol{v}$ (described as being generated by a DMC fed by $\boldsymbol{u}$) is available at the decoder.

The stochastic likelihood decoder estimates $\boldsymbol{u}$, using the channel output $\boldsymbol{y} = (y_1, \ldots, y_n)$ and the SI vector $\boldsymbol{v} = (v_1, \ldots, v_n)$, according to

$$\Pr\{\hat{u} = \boldsymbol{u}_0 | \boldsymbol{v}, \boldsymbol{y}\} = \frac{P(\boldsymbol{u}_0, \boldsymbol{v}) W(\boldsymbol{y}|\boldsymbol{x}[\boldsymbol{u}_0])}{\sum_{\boldsymbol{u}} P(\boldsymbol{u}, \boldsymbol{v}) W(\boldsymbol{y}|\boldsymbol{x}[\boldsymbol{u}])}. \tag{28}$$

Accordingly, let us define the GLD for the source–channel coding system as

$$\Pr\{\hat{u} = \boldsymbol{u}_0 | \boldsymbol{v}, \boldsymbol{y}\} = \frac{\exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u}_0)\boldsymbol{y}})]\}}{\sum_{\boldsymbol{u}} \exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u})\boldsymbol{y}})]\}}, \tag{29}$$

where g is as before and similarly, f is a continuous function of the joint empirical distribution of $\boldsymbol{u}$ and $\boldsymbol{v}$, $\hat{P}_{\boldsymbol{u}\boldsymbol{v}}$. The average probability of error of the GLD in this setting is taken w.r.t. the joint ensemble of the random binning codes and the random channel codes described above. We refer to the asymptotic exponential rate of this average error probability as the *random binning–coding error exponent* of the GLD.

In order to characterize the random binning–coding error exponent of this GLD, we define the following functions. For given joint distributions $Q_{U'V}$ and $Q_{X'Y}$ of the pairs of random variables (U', V) and (X', Y), respectively, we first define

$$h(Q_{U'V}, Q_{X'Y}) = f(Q_{U'V}) + g(Q_{X'Y}). \tag{30}$$

Next define

$$E_1(R, Q_{UV}) = \min_{Q_{U'V}} [[f(Q_{UV}) - f(Q_{U'V})]_+ + R - H(U'|V)]_+, \tag{31}$$

where $H(U'|V)$ is the conditional entropy of U' given V induced by $Q_{U'V}$, and

$$E_2(R) = \min_{Q_{UV}} \{D(Q_{UV} \| P_{UV}) + E_1(R, Q_{UV})\}. \tag{32}$$

Now, for given joint distributions Q_{UV}, Q_{XY}, $Q_{U'V}$ and $Q_{X'Y}$, define

$$E_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}) = [[h(Q_{UV}, Q_{XY}) - h(Q_{U'V}, Q_{X'Y})]_+ + I(X';Y) - H(U'|V)]_+, \tag{33}$$

where $I(X';Y)$ is the mutual information between X' given Y induced by $Q_{X'Y}$, and

$$E_4(Q_{UV}, Q_{XY}) \;\; = \;\; \min_{Q_{U'V}, Q_{X'Y}} E_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}). \tag{34}$$

Finally, define

$$E_5 = \min_{Q_{UV}, Q_{XY}} [D(Q_{UV} \| P_{UV}) + D(Q_{Y|X} \| W | Q_X) + E_4(Q_{UV}, Q_{XY})]. \tag{35}$$

The following theorem is proved in Appendix A.

Theorem 1 *The random binning–coding error exponent of the GLD (29) is given by*

$$E(R) = \min\{E_2(R), E_5\}. \tag{36}$$

Discussion

The term $E_2(R)$ corresponds to an error that occurs in the source coding stage, namely, in the random binning. It is associated with confusion of the true source vector $\boldsymbol{u}$ with another possible source vector $\boldsymbol{u}'$, which is assigned to the same bin, that is, $b(\boldsymbol{u}') = b(\boldsymbol{u})$. The other term stems from the channel coding part. Here, the terms E_3 and E_4 play roles that are parallel to those of E_2 and E_3 of Section 3. In other words, every conditional type of $\{\boldsymbol{u}'\}$ given $\boldsymbol{v}$ can thought of as a message set that is effectively mapped into a channel sub-code at rate $H(U'|V)$, which is the exponential rate of the cardinality of a conditional type class. This conditional type of source vectors competes with the true source vector $\boldsymbol{u}$. When the binning rate R is small, the source coding exponent $E_2(R)$ dominates, as the low binning rate is the primary obstacle to reliable communication, not the channel noise. In the other extreme, when R is very large, the binning encoder becomes a one–to–one mapping (with high probability) and we actually pass from separate source- and channel coding to joint source–channel coding. Consequently, the dependence on R disappears.

The system considered in this section was also studied in [11], in the context of universal decoding, with the motivation that it provides a common umbrella to many relevant special cases, including: separate/joint source–channel coding with/without side information, pure source coding with decoder side information (Slepian–Wolf model), pure channel coding, and systematic coding (see motivating discussion in [11]). The generality of the functions f and g in (29) adds considerably many additional degrees of freedom to the model discussed, in each of the above mentioned special cases. The various interesting choices of g have already been discussed before. Parallel choices can be considered also for f, e.g., $f(Q) = \beta \boldsymbol{E}_Q \log P'(U,V)$ for a mismatched source metric, $f(Q) = -\beta H_Q(U|V)$ for a stochastic version of the universal minimum conditional entropy

decoder, and so on. In particular, the choice $f(Q)+g(Q')=\beta[I_{Q'}(X;Y)-H_Q(U|V)]$ is associated with a stochastic version of the universal source–channel decoder considered in [11] (which is in turn an extension of the one in [1]). It is not difficult to verify that the universal stochastic decoder (29), with this choice of $f(Q)+g(Q')$, achieves the random binning–coding error exponent of the optimal MAP decoder, for every $\beta \geq 1$. This extends the main result of [11], which associated with the corresponding deterministic decoder ($\beta \to \infty$).

5 Expurgated Bound

In this section, we return to the pure channel coding setting of Section 3 and derive an expurgated bound on the error probability of the GLD. For a given code $\mathcal{C}_n$, the probability of error given that message m was transmitted is given by

$$P_{\mathrm{e}|m}(\mathcal{C}_n) = \sum_{m' \neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}. \tag{37}$$

In order to characterize the expurgated exponent, we define first a few quantities. Let

$$\alpha(R, Q_Y) = \sup_{\{Q_{X|Y}:\ I(Q_{XY}) \leq R\}} [g(Q_{XY}) - I(Q_{XY})] + R, \tag{38}$$

and

$$\begin{aligned}
\Gamma(Q_{XX'}, R) &= \inf_{Q_{Y|XX'}} \big\{ D(Q_{Y|X}\|W|Q_X) + I_Q(X';Y|X) + \\
&\qquad [\max\{g(Q_{XY}), \alpha(R, Q_Y)\} - g(Q_{X'Y})]_+ \} \qquad (39) \\
&\equiv \inf_{Q_{Y|XX'}} \big\{ \boldsymbol{E}_Q \log[1/W(Y|X)] - H(Y|X, X') + \\
&\qquad [\max\{g(Q_{XY}), \alpha(R, Q_Y)\} - g(Q_{X'Y})]_+ \} \qquad (40)
\end{aligned}$$

Our main result in this section is the following.

Theorem 2 *There exists a sequence of constant composition codes, $\{\mathcal{C}_n,\ n = 1, 2, \ldots\}$, with composition Q_X, such that*

$$\liminf_{n\to\infty} \left[-\frac{\log P_{e|m}(\mathcal{C}_n)}{n} \right] \geq E_{ex}^{gld}(R, Q_X), \tag{41}$$

where

$$E_{ex}^{gld}(R, Q_X) = \inf_{\{Q_{XX'}:\ I_Q(X;X') \leq R,\ Q_{X'} = Q_X\}} [\Gamma(Q_{XX'}, R) + I_Q(X; X')] - R. \tag{42}$$

Note that the expression of eq. (42) has the same structure as the Csiszár–Körner-Marton (CKM) expurgated bound [3], [2], except that here the functional $\Gamma(Q_{XX'}, R)$ replaces the expected Bhattacharyya distance (under $Q_{XX'}$) that appears in the CKM expurgated bound. The difference, however, is that unlike the expected Bhattacharyya distance, $\Gamma(Q_{XX'}, R)$ depends, in general, on R. As a consequence, the behavior of $E_{\mathrm{ex}}^{\mathrm{gld}}(R, Q_X)$ at high rates is not necessarily affine as the CKM expurgated exponent, We will return to this point later on.

Proof of Theorem 2. Consider first the expression

$$Z_m(\boldsymbol{y}) \stackrel{\Delta}{=} \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}. \tag{43}$$

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

$$\mathcal{B}_\epsilon(m, \boldsymbol{y}) = \Big\{\mathcal{C}_n : \ Z_m(\boldsymbol{y}) \le \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}\Big\}. \tag{44}$$

In Appendix B, we show that the vast majority of constant composition codes $\{\mathcal{C}_n\}$ (whose composition is Q_X), are outside $\mathcal{B}_\epsilon(m, \boldsymbol{y})$, simultaneously for all m and all $\boldsymbol{y}$. More precisely, it is shown in Appendix B that, considering the ensemble of randomly selected constant codes of type Q_X,

$$\Pr\{\mathcal{B}_\epsilon(m, \boldsymbol{y})\} \le \exp\{-e^{n\epsilon} + n\epsilon + 1\}, \tag{45}$$

for every m and $\boldsymbol{y}$, and so, by the union bound, this means that

$$\Pr\left\{\bigcup_m \bigcup_{\boldsymbol{y}\in\mathcal{Y}^n} \mathcal{B}_\epsilon(m, \boldsymbol{y})\right\} \stackrel{\Delta}{=} \Pr\{\mathcal{B}_\epsilon\} \le e^{nR}|\mathcal{Y}|^n \exp\{-e^{n\epsilon} + n\epsilon + 1\}, \tag{46}$$

which still decays double–exponentially. Thus, for all codes in $\mathcal{G}_\epsilon = \mathcal{B}_\epsilon^c$, which is the vast majority of constant composition codes codes $\{\mathcal{C}_n\}$ with composition Q_X, we have $Z_m(\boldsymbol{y}) \ge \exp\{n\alpha(R-\epsilon, \hat{Q}_{\boldsymbol{y}})\}$ simultaneously for all $m = 0, 1, \ldots, M-1$ and $\boldsymbol{y} \in \mathcal{Y}^n$. Now, trivially,

$$\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \sum_{m'\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}} \le 1, \tag{47}$$

and for a code in $\mathcal{G}_\epsilon \stackrel{\Delta}{=} \mathcal{B}_\epsilon^c$, we also have

$$\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \sum_{m'\neq m}\exp\{ng(\hat{Q}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}} \le \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}}. \tag{48}$$

Thus, for such a code

$$\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \sum_{m'\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}} \le \min\left\{1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}}\right\}. \tag{49}$$

It follows that for every $\mathcal{C}_n \in \mathcal{G}_\epsilon$,

$$\begin{aligned}
P_{e|m}(\mathcal{C}_n) &\le \sum_{m'\neq m}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \min\left\{1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}}\right\} \\
&\doteq \sum_{m'\neq m}\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\exp\{-n[\max\{g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}), \alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\} - g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})]_+\} \\
&\doteq \sum_{m'\neq m}\exp\{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{x}_{m'}}, R-\epsilon)\} \\
&= \sum_{Q_{X'|X}:\ Q_{X'}=Q_X} N_m(Q_{XX'})\exp\{-n\Gamma(Q_{XX'}, R-\epsilon)\}
\end{aligned} \tag{50}$$

where $N_m(Q_{XX'})$ is the number of codewords $\{\boldsymbol{x}_{m'}\}$ whose joint type with $\boldsymbol{x}_m$ is exactly $Q_{XX'}$, and where

$$\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m)\exp\{-n[\max\{g(\hat{Q}_{\boldsymbol{x}_m\boldsymbol{y}}), \alpha(R-\epsilon, \hat{Q}_{\boldsymbol{y}})\} - g(\hat{Q}_{\boldsymbol{x}_{m'}\boldsymbol{y}})]_+\}$$

$$
\begin{aligned}
&\doteq \max_{Q_{Y|XX'}} \exp\{n(H_Q(Y|X,X') - H_Q(Y|X) - D(\hat{Q}_{Y|X}\|W|P) - \\
&\quad [\max\{g(Q_{XY}), \alpha(R-\epsilon, Q_Y)\} - g(Q_{X'Y})]_+)\} \\
&= \exp\left\{-n \min_{Q_{Y|XX'}} \Big[D(\hat{Q}_{Y|X}\|W|P) + I_Q(X';Y|X) + \right. \\
&\quad \left. [\max\{g(Q_{XY}), \alpha(R-\epsilon, Q_Y)\} - g(Q_{X'Y})]_+]\right\} \\
&= \exp\{-n\Gamma(Q_{XX'}, R-\epsilon)\}. \qquad (51)
\end{aligned}
$$

Now, as is shown in Appendix C, for most codes in $\mathcal{G}_\epsilon$,

$$
N_m(Q_{XX'}) \le \begin{cases} \exp\{n[R - I_Q(X;X')]\} & R \ge I_Q(X;X') \\ 0 & R < I_Q(X;X') \end{cases} \qquad (52)
$$

for all m and all $Q_{XX'}$, and so, considering the arbitrariness of ϵ, the expurgated error exponent is given by

$$
E_{\rm ex}^{\rm gld}(R, Q_X) = \min_{\{Q_{XX'}:\ I_Q(X;X')\le, Q_{X'}=Q_X\}} [\Gamma(Q_{XX'}, R) + I_Q(X;X')] - R \qquad (53)
$$

This completes the proof of Theorem 2. $\square$

It is interesting to note an important difference between the first steps in the derivation in the proof of Theorem 2 above, and the first steps in the derivation of the ordinary expurgated bound. While for the ordinary expurgated bound, the starting point is the inequality

$$
P_{e|m}(\mathcal{C}_n) = \sum_{m'\neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \sqrt{\frac{W(\boldsymbol{y}|\boldsymbol{x}_{m'})}{W(\boldsymbol{y}|\boldsymbol{x}_m)}} \qquad (54)
$$

or, more generally,

$$
P_{e|m}(\mathcal{C}_n) = \sum_{m'\neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \left[\frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})}}\right]^\gamma, \quad \gamma \ge 0, \qquad (55)
$$

the above derivation in the proof of Theorem 2 begins from from the inequality

$$
P_{e|m}(\mathcal{C}_n) = \sum_{m'\neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \min\left\{1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}}\right\}. \qquad (56)
$$

It is easy to argue that for $\gamma \in [0,1]$ (and in particular, $\gamma = 1/2$, used at least when $g(Q) = \sum_{x,y} Q(x,y)\ln W(y|x)$):

$$
\min\left\{1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}}\right\} \le \left[\frac{e^{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})}}\right]^\gamma. \qquad (57)
$$

To see why this is true, let us distinguish between the cases $g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}}) \le g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})$ and $g(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}}) > g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})$. In the former case,

$$
\begin{aligned}
&\min\left\{1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}}\right\} \qquad (58) \\
&\le \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}} \qquad (59)
\end{aligned}
$$

$$\leq \left[\frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}}\right]^{\gamma}. \tag{60}$$

In the latter case, the right–hand side of (57) exceeds unity, whereas the left–hand side is always less than unity. Since all the subsequent derivations in the proof of Theorem 2 are exponentially tight (by the method of types), the conclusion from this observation is that at least for the choice $g(Q) = \sum_{x,y} Q(x,y) \log W(y|x)$, the new expurgated bound, $E_{\rm ex}^{\rm gld}(R, Q_X)$, is *at least as tight* as the CKM expurgated bound. In the next example, we demonstrate that it may indeed be strictly tighter than the CKM expurgated bound at least at relatively high rates.

Example – the Z-Channel. Consider the z–channel with $\mathcal{X} = \mathcal{Y} = \{0,1\}$, which is parametrized by $w \in [0,1]$ as follows:

$$W(y|x) = \begin{cases} w & x = y = 0 \\ 1-w & x = 0,\ y = 1 \\ 0 & x = 1,\ y = 0 \\ 1 & x = y = 1 \end{cases} \tag{61}$$

and let the input assignment be $Q_X(0) = Q_X(1) = 1/2$. Let $g(Q) = \boldsymbol{E}_Q \log W(Y|X)$. In the case of the z–channel, any joint empirical distribution Q_{XY} for which $g(Q_{XY}) > -\infty$, must also be of the z–form:

$$Q_{XY}(x,y) = \begin{cases} q/2 & x = y = 0 \\ (1-q)/2 & x = 0,\ y = 1 \\ 0 & x = 1,\ y = 0 \\ 1/2 & x = y = 1 \end{cases} \tag{62}$$

where $q \in [0,1]$ designates the associated empirical transition probability from $X = 0$ to $Y = 0$. Thus,

$$Q_Y(y) = \begin{cases} q/2 & y = 0 \\ 1 - q/2 & y = 1 \end{cases} \tag{63}$$

Now,

$$g(Q_{XY}) = g(q) \stackrel{\Delta}{=} \begin{cases} \frac{q}{2}\log w + \frac{1-q}{2}\log(1-w) & Q_{XY}(0,1) = 0 \\ -\infty & Q_{XY}(0,1) > 0 \end{cases} \tag{64}$$

We begin from the calculation of $\alpha(R, Q_Y)$, which will be denoted by $\alpha(R,q)$. We observe that for a given Q_Y, which means actually, a given q, there is only one empirical channel, so here, the set $\{Q_{X|Y} : I(Q_{XY}) \leq R\}$ is either a singleton or an empty set, depending on q and R. The mutual information for a given q is

$$I(Q_{XY}) = I(q) = h\left(\frac{q}{2}\right) - \frac{1}{2}h(q), \tag{65}$$

where $h(\cdot)$ is the binary entropy function. Thus,

$$\alpha(R,q) = \begin{cases} \frac{q}{2}\log w + \frac{1-q}{2}\log(1-w) - I(q) + R & I(q) \leq R \\ -\infty & I(q) > R \end{cases} \tag{66}$$

and so,

$$\max\{g(Q_{XY}), \alpha(R, Q_Y)\} = \max\{g(q), \alpha(R,q)\} = g(q) + [R - I(q)]_+. \tag{67}$$

which yields

$$[\max\{g(Q_{XY}), \alpha(R, Q_Y)\} - g(Q_{X'Y})]_+ = [g(q) + [R - I(q)]_+ - g(q)]_+ = [R - I(q)]_+. \tag{68}$$

For a given q, which is actually a given $\hat{P}_{\boldsymbol{y}}$, and a given pair of codewords $\{\boldsymbol{x}_m, \boldsymbol{x}_{m'}\}$, with a joint empirical distribution $\hat{Q}_{XX'}$. we are summing in (51) the expression $e^{ng(q)} \cdot e^{-n[R-I(q)]_+}$ over all $\boldsymbol{y}$, but the summand is positive only for $\boldsymbol{y}$ for which both $\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}}$ and $\hat{P}_{\boldsymbol{x}_{m'} \boldsymbol{y}}$ agree with Q_{XY} as defined above (with q). This can be the case only if $q \leq 2Q_{XX'}(0,0)$ and

$$Q_{Y|XX'}(0|0,0) = \frac{q}{2Q_{XX'}(0,0)} \tag{69}$$

$$Q_{Y|XX'}(1|0,0) = 1 - \frac{q}{2Q_{XX'}(0,0)} \tag{70}$$

$$Q_{Y|XX'}(0|0,1) = Q_{Y|XX'}(0|1,0) = Q_{Y|XX'}(0|1,1) = 0. \tag{71}$$

The above–mentioned sum is therefore of the exponential order of

$$\exp\left\{n\left[Q_{XX'}(0,0)h_2\left(\frac{q}{2Q_{XX'}(0,0)}\right) + g(q) - [R-I(q)]_+\right]\right\}.$$

and so,

$$\Gamma(Q_{XX'}, R) = [R-I(q)]_+ - g(q) - Q_{XX'}(0,0) \cdot h\left(\frac{q}{2Q_{XX'}(0,0)}\right). \tag{72}$$

Let us denote $\theta = Q_{XX'}(0,0)$ ($\theta \leq 1/2$), so

$$\Gamma(\theta) = [R-I(q)]_+ - g(q) - \theta h\left(\frac{q}{2\theta}\right). \tag{73}$$

Note that since both marginals of $Q_{XX'}$ are binary symmetric sources, then $Q_{XX'}(1,0) = Q_{XX'}(0,1) = 1/2 - \theta$ and $Q_{XX'}(1,1) = \theta$. Now,

$$I(Q_{XX'}) = I(\theta) \stackrel{\Delta}{=} 2\theta \log\frac{\theta}{1/4} + 2\left(\frac{1}{2} - \theta\right)\log\frac{1/2-\theta}{1/4} = \log 2 - h(2\theta). \tag{74}$$

It follows that $Q_{XX'}(1,0) = Q_{XX'}(0,1) = 1/2 - \theta$ and $Q_{XX'}(1,1) = \theta$. Now,

$$I(Q_{XX'}) = I(\theta) \stackrel{\Delta}{=} 2\theta \log\frac{\theta}{1/4} + 2\left(\frac{1}{2} - \theta\right)\log\frac{1/2-\theta}{1/4} = \log 2 - h(2\theta). \tag{75}$$

It follows that

$$E_{\rm ex}^{\rm gld}(R, Q_X) = \min_{\{\theta:\ \log 2 - h(2\theta) \leq R,\ \theta \leq 1/2\}} \min_{q \leq 2\theta} \left\{[R-I(q)]_+ - g(q) - h(2\theta) - \theta h\left(\frac{q}{2\theta}\right)\right\} + \log 2 - R. \tag{76}$$

The ordinary expurgated bound (CKM), on the other hand, is given by

$$E_{\rm ex}(R, Q_X) = \begin{cases} -\frac{1}{2}h^{-1}(\log 2 - R)\log(1-w) & R \leq \log 2 - h\left(\frac{1}{1+\sqrt{1-w}}\right) \\ \log\frac{2}{1+\sqrt{1-w}} - R & R > \log 2 - h\left(\frac{1}{1+\sqrt{1-w}}\right) \end{cases} \tag{77}$$

Interestingly, if the non–negative term $[R-I(q)]_+$ is discarded from (76), which results in a lower bound to $E_{\rm ex}^{\rm gld}(R, Q_X)$, then the minimization of the remaining expression can easily be carried out analytically, and it turns out to yield exactly the same expression as that of the CKM expurgated exponent in eq. (77). Thus, it is the term $[R-I(q)]_+$ that has the potential to improve on the CKM expurgated exponent, at least at relatively high rates. Indeed, in Fig. 2, we see comparative plots of the CKM expurgated exponent (in blue) and the new expurgated exponent (in green) as well as the random coding error exponent (in red), all for $w = 0.9$. As can be seen, while the CKM expurgated exponent descends linearly for high rates (as is well known), the new expurgated

exponent departs from it in the high rate region, and it seems to follow the curve of the random coding exponent. In other words, at least in this example, the new expurgated exponent seems to follow the maximum between the random coding exponent and the CKM exponent, and therefore to improve on the CKM expurgated exponent at high rates. This is in spite of the fact that the new expurgated exponent was developed for a sub-optimal decoder. We believe that one of the reasons for this improvement is that the new expurgated bound is not based on the union bound, which is inherently the starting point of the classic expurgated bound, and also its weakness at high rates.

In future research, it would be interesting to explore the new expurgated bound in additional examples and see if it may improve on existing lower bounds to the reliability function, and thereby shrink the gap between the well known lower bounds and upper bounds to this function.

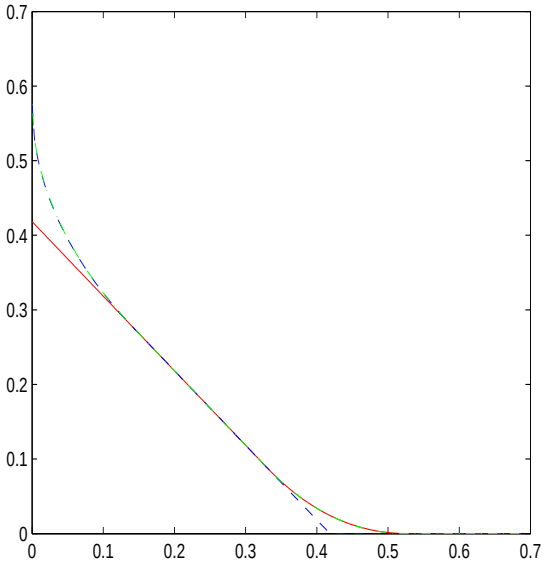

Figure 2: Various exponents for the z–channel with parameter $w = 0.9$. The red (solid) curve is the random coding bound, the blue (dashed) one is the classical expurgated bound, and the green (dashed) curve is the new expurgated bound. The latter seems to behave as the maximum between the first two.

Appendix A

Proof of Theorem 1. The proof is based on the same technique as in the derivation in Section 3, as well as in [11]. The probability of error is given by

$$\bar{P}_e = \boldsymbol{E}\left\{\frac{\sum_{\boldsymbol{u}'\neq\boldsymbol{U}}\exp\{n[f(\hat{Q}_{\boldsymbol{u}'\boldsymbol{V}})+g(\hat{Q}_{\boldsymbol{X}(\boldsymbol{u}')\boldsymbol{Y}})]\}}{\sum_{\boldsymbol{u}'}\exp\{n[f(\hat{Q}_{\boldsymbol{u}'\boldsymbol{V}})+g(\hat{Q}_{\boldsymbol{X}(\boldsymbol{u}')\boldsymbol{Y}})]\}}\right\}. \tag{A.1}$$

Let us condition first on $(\boldsymbol{U}=\boldsymbol{u}_0, \boldsymbol{V}=\boldsymbol{v}, b(\boldsymbol{u}_0)=j_0, \boldsymbol{X}(j_0)=\boldsymbol{x}_0, \boldsymbol{Y}=\boldsymbol{y})$ and take the expectation only w.r.t. the random binning of source vectors other than $\boldsymbol{u}_0$ and codewords other than $\boldsymbol{X}(j_0)$. Using the same technique as before, we assess the conditional probability of error as

$$\bar{P}_{\mathrm{e}}(\boldsymbol{u}_0, \boldsymbol{v}, j_0, \boldsymbol{x}_0, \boldsymbol{y})$$

$$\dot{=} \int_0^\infty e^{-n\theta} \cdot \Pr\left\{\sum_{\boldsymbol{u}'} \exp\{n[f(\hat{Q}_{\boldsymbol{u}'\boldsymbol{v}}) + g(\hat{Q}_{\boldsymbol{X}(\boldsymbol{u}')\boldsymbol{y}})]\} \geq \right.$$
$$\left. \exp\{n[f(\hat{Q}_{\boldsymbol{u}_0\boldsymbol{v}}) + g(\hat{Q}_{\boldsymbol{X}(\boldsymbol{u}_0)\boldsymbol{y}}) - \theta]\}\right\} \mathrm{d}\theta. \tag{A.2}$$

We first condition on the channel code $\mathcal{C}_n = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$, $M = e^{nR}$, and calculate the probability only w.r.t. the randomness of the binning. Consider the following decomposition:

$$\begin{aligned}
&\sum_{\boldsymbol{u}'} \exp\{n[f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u}')\boldsymbol{y}})]\} \\
&= \sum_{\boldsymbol{u}':b(\boldsymbol{u}')=b(\boldsymbol{u})} \exp\{n[f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u}')\boldsymbol{y}})]\} + \sum_{\boldsymbol{u}':b(\boldsymbol{u}')\neq b(\boldsymbol{u})} \exp\{n[f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u}')\boldsymbol{y}})]\} \\
&= \exp\{ng(\hat{P}_{\boldsymbol{x}(\boldsymbol{u})\boldsymbol{y}})\} \cdot \sum_{\boldsymbol{u}':b(\boldsymbol{u}')=b(\boldsymbol{u})} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\} + \sum_{\boldsymbol{u}':b(\boldsymbol{u}')\neq b(\boldsymbol{u})} \exp\{n[f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u}')\boldsymbol{y}})]\} \\
&\stackrel{\Delta}{=} Z_1 + Z_2.
\end{aligned} \tag{A.3}$$

Then, obviously,

$$\begin{aligned}
&\Pr\left\{\sum_{\boldsymbol{u}'} \exp\{n[f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u}')\boldsymbol{y}})]\} \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}}) - \theta]\}\right\} \\
\dot{=}\ &\Pr\left\{Z_1 \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}}) - \theta]\}\right\} + \\
&\Pr\left\{Z_2 \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}}) - \theta]\}\right\}.
\end{aligned} \tag{A.4}$$

Let us begin with the first term,

$$\begin{aligned}
&\Pr\left\{Z_1 \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u})\boldsymbol{y}}) - \theta]\}\right\} \\
=\ &\Pr\left\{\sum_{\boldsymbol{u}':b(\boldsymbol{u}')=b(\boldsymbol{u})} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\} \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - \theta]\}\right\}.
\end{aligned} \tag{A.5}$$

Denote $\mathcal{C}_n(\boldsymbol{u}) = \{\boldsymbol{u}' :\ b(\boldsymbol{u}') = b(\boldsymbol{u})\}$, and for a given conditional type $Q_{U'|V}$ of $\boldsymbol{u}'$ given $\boldsymbol{v}$, let

$$N(Q_{U'|V}) = |\mathcal{C}_n(\boldsymbol{u}') \bigcap \mathcal{T}(Q_{U'|V}|\boldsymbol{v})|. \tag{A.6}$$

Obviously, $N(Q_{U'|V})$ is a binomial random variable with $|\mathcal{T}(Q_{U'|V}|\boldsymbol{v})| \dot{=} e^{nH(U'|V)}$ trials and probability of success e^{-nR}. Therefore,

$$\begin{aligned}
&\Pr\left\{\sum_{\boldsymbol{u}':\boldsymbol{x}(\boldsymbol{u}')=\boldsymbol{x}(\boldsymbol{u})} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\} \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - \theta]\}\right\} \\
=\ &\Pr\left\{\sum_{Q_{U'V}} N(Q_{U'V}) e^{nf(Q_{U'V})} \geq e^{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - \theta]}\right\} \\
\dot{=}\ &\max_{Q_{U'V}} \Pr\left\{N(Q_{U'V}) \geq e^{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V}) - \theta]}\right\} \\
=\ &\exp\{-nF_1(R, \hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}, \theta)\}
\end{aligned} \tag{A.7}$$

where

$$F_1(R, \hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}, \theta) = \min_{Q_{U'V}} \{[R - H(U'|V)]_+ :\ f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V}) - \theta \leq [H(U'|V) - R]_+\}$$

$$= \min_{Q_{U'V}} \{[R - H(U'|V)]_+ : \ \theta \geq f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V}) - [H(U'|V) - R]_+\} \quad \text{(A.8)}$$

where the minimum over $\{Q_{U'V}\}$ is subject to the constraint that its V–marginal coincides with $\hat{P}_{\boldsymbol{v}}$. Consequently, the contribution of Z_1 to the conditional probability of error, which we denote by $\bar{P}_{\mathrm{e}1}(\boldsymbol{u}_0, \boldsymbol{v}, j_0, \boldsymbol{x}_0, \boldsymbol{y})$, is the following:

$$\begin{aligned}
&\bar{P}_{\mathrm{e}1}(\boldsymbol{u}_0, \boldsymbol{v}, j_0, \boldsymbol{x}_0, \boldsymbol{y}) \\
&\doteq \int_0^\infty e^{-n\theta} \exp\{-nF_1(R, \hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}, \theta)\} \mathrm{d}\theta \\
&\doteq \int_{[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V}) - [H(U'|V) - R]_+]_+}^\infty e^{-n\theta} \exp\{-n[R - H(U'|V)]_+\} \mathrm{d}\theta \\
&\doteq e^{-nE_1(R, \hat{P}_{\boldsymbol{u}_0\boldsymbol{v}})}
\end{aligned} \quad \text{(A.9)}$$

with

$$\begin{aligned}
E_1(R, \hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) &= \min_{Q_{U'V}} \Big\{[R - H(U'|V)]_+ + [f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V}) - [H(U'|V) - R]_+]_+\Big\} \\
&= \min_{Q_{U'V}} \begin{cases} [f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V}) + R - H(U'|V)]_+ & R < H(U'|V) \\ R - H(U'|V) + [f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V})]_+ & R \geq H(U'|V) \end{cases} \\
&= \min_{Q_{U'V}} [[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) - f(Q_{U'V})]_+ + R - H(U'|V)]_+
\end{aligned} \quad \text{(A.10)}$$

The overall contribution of Z_1 to the (unconditional) probability of error is therefore

$$\bar{P}_{\mathrm{e}1} \doteq e^{-nE_2(R)} \quad \text{(A.11)}$$

where

$$E_2(R) = \min_{Q_{UV}} \{D(Q_{UV} \| P_{UV}) + E_1(R, Q_{UV})\}, \quad \text{(A.12)}$$

as defined also in Section 4.

We next move on to handle Z_2, which we have defined as

$$\begin{aligned}
Z_2 &= \sum_{\boldsymbol{u}' :\ b(\boldsymbol{u}') \neq b(\boldsymbol{u})} \exp\{n[f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}(\boldsymbol{u}')\boldsymbol{y}})]\} \\
&= \sum_{\mathcal{T}(Q_{U'|V}|\boldsymbol{v})} e^{nf(Q_{U'V})} \sum_{\mathcal{T}(Q_{X'|Y}|\boldsymbol{y})} e^{ng(Q_{X'Y})} \sum_{\boldsymbol{u}' \in \mathcal{T}(Q_{U'|V}|\boldsymbol{v})} \mathcal{I}[b(\boldsymbol{u}') \neq b(\boldsymbol{u})] \cdot \mathcal{I}[\boldsymbol{X}(\boldsymbol{u}') \in \mathcal{T}(Q_{X'|Y}|\boldsymbol{y})] \\
&\triangleq \sum_{\mathcal{T}(Q_{U'|V}|\boldsymbol{v})} e^{nf(Q_{U'V})} \sum_{\mathcal{T}(Q_{X'|Y}|\boldsymbol{y})} e^{ng(Q_{X'Y})} N(Q_{U'V}, Q_{X'Y}).
\end{aligned} \quad \text{(A.13)}$$

Now, for a given channel code $\mathcal{C}_n$, $N(Q_{U'V}, Q_{X'Y})$ is a binomial random variable with exponentially $e^{nH(U'|V)}$ trials and probability of success $(1 - e^{-nR})|(\mathcal{C}_n \setminus \{\boldsymbol{x}_0\}) \cap \mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})|/(|\mathcal{C}_n| - 1) \doteq e^{-nR}|(\mathcal{C}_n \setminus \{\boldsymbol{x}_0\}) \cap \mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})| \triangleq e^{-nS(Q_{X'Y})}$. Thus,

$$\begin{aligned}
&\Pr\left\{\sum_{\mathcal{T}(Q_{U'|V}|\boldsymbol{v})} e^{nf(Q_{U'V})} \sum_{\mathcal{T}(Q_{X'|Y}|\boldsymbol{y})} e^{ng(Q_{X'Y})} N(Q_{U'V}, Q_{X'Y}) \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}}) - \theta]\}\right\} \\
&\doteq \max_{Q_{U'V}, Q_{X'Y}} \Pr\Big\{N(Q_{U'V}, Q_{X'Y}) \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}_0\boldsymbol{v}}) + g(\hat{P}_{\boldsymbol{x}_0\boldsymbol{y}}) - f(Q_{U'V}) - g(Q_{X'Y}) - \theta]\}\Big\}.
\end{aligned} \quad \text{(A.14)}$$

We henceforth use the shorthand notation $h(Q_{U'V}, Q_{X'Y}) = f(Q_{U'V}) + g(Q_{X'Y})$, as defined in Section 4. The exponential order of the last probability is given by

$$\begin{aligned} F_2(Q_{UV}, Q_{XY}, \theta) &= \min_{Q_{U'V}, Q_{X'Y}} \big\{[S(Q_{X'Y}) - H(U'|V)]_+ : \ \theta \geq h(Q_{UV}, Q_{XY}) - \\ &\qquad h(Q_{U'V}, Q_{X'Y}) - [H(U'|V) - S(Q_{X'Y})]_+\big\} \end{aligned} \tag{A.15}$$

Now,

$$\begin{aligned} &\max_{Q_{U'V}, Q_{X'Y}} \int_0^\infty e^{-n\theta} \exp\{-nF_2(Q_{UV}, Q_{XY}, \theta)\} \mathrm{d}\theta \\ \doteq\ &\max_{Q_{U'V}, Q_{X'Y}} \int_{[h(Q_{UV},(Q_{XY}) - h(Q_{U'V},(Q_{X'Y}) - [H(U'|V) - S(Q_{X'Y})]_+]_+}^\infty e^{-n\theta} \exp\{-n[S(Q_{X'Y}) - H(U'|V)]_+\} \mathrm{d}\theta \\ \doteq\ &\exp\{-n \min_{Q_{U'V}, Q_{X'Y}} E_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, S(Q_{X'Y}))\} \end{aligned} \tag{A.16}$$

where for a given S,

$$\begin{aligned} &E_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, S) \\ \stackrel{\Delta}{=}\ &\big\{[S - H(U'|V)]_+ + [h(Q_{UV}, Q_{XY}) - h(Q_{U'V}, Q_{X'Y}) - [H(U'|V) - S]_+]_+\big\} \\ =\ &\min_{Q_{U'V}, Q_{X'Y}} \begin{cases} [h(Q_{UV}, Q_{XY}) - h(Q_{U'V}, Q_{X'Y}) + S - H(U'|V)]_+ & S < H(U'|V) \\ S - H(U'|V) + [h(Q_{UV}, Q_{XY}) - h(Q_{U'V}, Q_{X'Y})]_+ & S \geq H(U'|V) \end{cases} \\ =\ &[[h(Q_{UV}, Q_{XY}) - h(Q_{U'V}, Q_{X'Y})]_+ + S - H(U'|V)]_+ \end{aligned} \tag{A.17}$$

It remains to average this expression w.r.t. the randomness of the channel code $\mathcal{C}_n$. For a given $Q_{X'Y}$, it follows from the definition of $S(Q_{X'Y})$ that $S(Q_{X'Y}) = R - \frac{1}{n}\log N(Q_{X'Y})$ where $N(Q_{X'Y})$ is the number of codewords in $\mathcal{C}_n \setminus \{\boldsymbol{x}_0\}$ whose joint type with $\boldsymbol{y}$ is $Q_{X'Y}$. Now, $N(Q_{X'Y})$ is a binomial random variable with e^{nR} trials and probability of success of the exponential order of $e^{-nI(X';Y)}$. Therefore, the expectation of $\exp\{-nE_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, S(Q_{X'Y}))\}$ w.r.t. the randomness of the code is assessed as follows. Let $\epsilon > 0$ be arbitrarily small. Then the desired average is upper bounded by

$$\begin{aligned} &\sum_{i \geq 0} \Pr\left\{e^{ni\epsilon} \leq N(Q_{X'Y}) < e^{n(i+1)\epsilon}\right\} \cdot \exp\{-nE_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, R - (i+1)\epsilon)\} \\ \doteq\ &\max_{0 \leq i \leq [R - I(X';Y)]_+/\epsilon} \exp\{-n[I(X';Y) - R]_+\} \cdot \exp\{-nE_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, R - (i+1)\epsilon)\} \\ \doteq\ &\exp\{-n[I(X';Y) - R]_+\} \cdot \exp\{-nE_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, R - [R - I(X';Y)]_+) - \epsilon\} \\ =\ &\exp\{-nE_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, I(X';Y) - \epsilon)\}, \end{aligned} \tag{A.18}$$

but since $\epsilon > 0$ is arbitrary, $E_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, I(X';Y))$ can be approached as closely as desired. We henceforth omit the term ϵ in E_3 and denote

$$E_4(Q_{UV}, Q_{XY}) = \min_{Q_{U'V}, Q_{X'Y}} E_3(Q_{UV}, Q_{XY}, Q_{U'V}, Q_{X'Y}, I(X';Y)), \tag{A.19}$$

the overall contribution of Z_2 to the average probability of error is of the exponential order of e^{-nE_5}, where

$$E_5 = \min_{Q_{UV}, Q_{XY}} [D(Q_{UV} \| P_{UV}) + D(Q_{Y|X} \| W|P) + E_4(Q_{UV}, Q_{XY})]. \tag{A.20}$$

Finally, the overall exponent is

$$E(R) = \min\{E_2(R), E_5\}, \tag{A.21}$$

which completes the proof of Theorem 1.

Appendix B

Proof of Eq. (45).

We show that the vast majority of codes have $Z_m(\boldsymbol{y}) \geq \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}$ for all m and $\boldsymbol{y}$. First, observe that

$$Z_m(\boldsymbol{y}) = \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\} = \sum_{Q} N_{\boldsymbol{y}}(Q) e^{ng(Q)}. \tag{B.1}$$

Thus, considering the randomness of $\{\boldsymbol{X}_{m'}\}$,

$$\begin{aligned}
& \Pr\left\{Z_m(\boldsymbol{y}) \leq \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\
= \;& \Pr\left\{\sum_{Q} N_{\boldsymbol{y}}(Q) e^{ng(Q)} \leq \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\
\leq \;& \Pr\left\{\max_{Q} N_{\boldsymbol{y}}(Q) e^{ng(Q)} \leq \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\
= \;& \Pr\bigcap_{Q}\left\{N_{\boldsymbol{y}}(Q) e^{ng(Q)} \leq \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\
= \;& \Pr\bigcap_{Q}\left\{N_{\boldsymbol{y}}(Q) \leq \exp\{n[\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}) - g(Q)]\}\right\}.
\end{aligned} \tag{B.2}$$

Now, $N_{\boldsymbol{y}}(Q)$ is a binomial random variable with e^{nR} trials and success rate of the exponential order of $e^{-nI(Q)}$. We now argue that by the very definition of $\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})$, there must exist some $Q^*_{X|Y}$ such that for $Q^* = \hat{P}_{\boldsymbol{y}} \times Q^*_{X|Y}$, $I(Q^*) \leq R-\epsilon$ and $R-\epsilon-I(Q^*) \geq \alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}) - g(Q^*)$. To see why this is true, assume conversely, that for every $Q_{X|Y}$, which defines $Q = \hat{P}_{\boldsymbol{y}} \times Q_{X|Y}$, either $I(Q) > R-\epsilon$ or $R - I(Q) - \epsilon < \alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}) - g(Q)$, which means that for every Q

$$R - \epsilon < \max\{I(Q), I(Q) + \alpha(R-\epsilon, \hat{Q}_{\boldsymbol{y}}) - g(Q)\} = I(Q) + [\alpha(R-\epsilon, \hat{Q}_{\boldsymbol{y}}) - g(Q)]_+ \tag{B.3}$$

which implies in turn that for every $Q_{X|Y}$ there exists $t \in [0,1]$ such that

$$R - \epsilon < I(Q) + t[\alpha(R-\epsilon, \hat{Q}_{\boldsymbol{y}}) - g(Q)] \tag{B.4}$$

or equivalently,

$$\begin{aligned}
\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}) \;&>\; \max_{Q_{X|Y}} \min_{0 \leq t \leq 1} g(Q) + \frac{R - I(Q) - \epsilon}{t} \\
&= \max_{Q_{X|Y}} \begin{cases} g(Q) + R - I(Q) - \epsilon & I(Q) \leq R-\epsilon \\ -\infty & I(Q) > R-\epsilon \end{cases} \\
&= \max_{\{Q_{X|Y}:\; I(Q) \leq R-\epsilon\}} [g(Q) - I(Q)] + R - \epsilon \\
&\equiv \alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}),
\end{aligned} \tag{B.5}$$

which is a contradiction. Let then $Q^*_{X|Y}$ be as defined above. Then,

$$\Pr\bigcap_{Q}\left\{N_{\boldsymbol{y}}(Q) \leq \exp\{n[\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}) - g(Q)]\}\right\}$$

$$\leq \quad \Pr\left\{N_{\boldsymbol{y}}(Q^*) \leq \exp\{n[\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}) - g(Q^*)]\}\right\}. \tag{B.6}$$

Now, we know that $I(Q^*) \leq R - \epsilon$ and $R - I(Q^*) - \epsilon \geq \alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}}) - g(Q^*)$. By the Chernoff bound, the probability in question is upper bounded by

$$\exp\left\{-e^{nR} D(e^{-an} \| e^{-bn})\right\}, \tag{B.7}$$

where $a = R + g(Q^*) - \alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})$ and $b = I(Q^*)$. Noting that $a - b \geq \epsilon$, we can easily lower bound the binary divergence as follows (see [7, Section 6.3]):

$$\begin{aligned} D(e^{-an} \| e^{-bn}) &\geq e^{-bn}\{1 - e^{-(a-b)n}[1 + n(a-b)]\} \\ &\geq e^{-nI(Q^*)}[1 - e^{-n\epsilon}(1+n\epsilon)], \end{aligned} \tag{B.8}$$

where in the last passage, we have used the decreasing monotonicity of the function $f(t) = (1+t)e^{-t}$ for $t \geq 0$. Thus,

$$\begin{aligned} \Pr\left\{N_{\boldsymbol{y}}(Q^*) \leq \exp\{n[\alpha(R, \hat{P}_{\boldsymbol{y}}) - g(Q^*) - \epsilon]\}\right\} &\leq \exp\left\{-e^{nR} \cdot e^{-nI(Q)}[1 - e^{-n\epsilon}(1+n\epsilon)]\right\} \\ &\leq \exp\left\{-e^{n\epsilon}[1 - e^{-n\epsilon}(1+n\epsilon)]\right\} \\ &= \exp\left\{-e^{n\epsilon} + n\epsilon + 1\right\}, \end{aligned} \tag{B.9}$$

which completes the proof of eq. (45).

Appendix C

Proof of Eq. (52). The proof is very similar to the proof of Theorem 2 in [12], but there is a small twist due to the limitation to codes in $\mathcal{G}_\epsilon$ herein. Consider the uniform random selection of codebooks $\{\mathcal{C}_n\}$ in $\mathcal{G}_\epsilon$. For every given code $\mathcal{C}_n \in \mathcal{G}_\epsilon$ and a given message m, let $N_m(Q_{XX'}, \mathcal{C}_n)$ be the number of $\{m'\}$, all different from m, for which $(\boldsymbol{x}_m, \boldsymbol{x}_{m'})$ has a given joint empirical distribution $Q_{XX'}$ of a pair of random variables taking on values in $\mathcal{X}^2$, whose single–letter marginals (which are the individual empirical distributions of the various codewords) all coincide with Q_X. Our goal here is to show that for every $\epsilon > 0$ and sufficiently large n, there exists a code $\mathcal{C}_n \in \mathcal{G}_\epsilon$ of rate (essentially) R, that satisfies, for every message m and every $Q_{XX'}$, eq. (52), namely,

$$\begin{aligned} N_m(Q_{XX'}, \mathcal{C}_n) &\leq N^*(Q_{XX'}) \\ &\stackrel{\Delta}{=} \begin{cases} \exp\{n[R - I_Q(X;X') + \epsilon]\} & R \geq I_Q(X;X') - \epsilon \\ 0 & R < I_Q(X;X') - \epsilon \end{cases} \end{aligned} \tag{C.1}$$

To see why this is true, consider a random selection of the code $\mathcal{C}_n$ within $\mathcal{G}_\epsilon$. Then, obviously,

$$\overline{N(Q_{XX'})} \stackrel{\Delta}{=} \frac{1}{M}\sum_{m=0}^{M-1} \boldsymbol{E}\{N_m(Q_{XX'}, \mathcal{C}_n) | \mathcal{G}_\epsilon\} \tag{C.2}$$

$$\leq \frac{1}{M}\sum_{m=0}^{M-1} \frac{\boldsymbol{E}\{N_m(Q_{XX'}, \mathcal{C}_n)\}}{\Pr\{\mathcal{G}_\epsilon\}} \tag{C.3}$$

$$\doteq \boldsymbol{E}\{N_0(Q_{XX'}, \mathcal{C}_n)\} \tag{C.4}$$

$$= M \cdot \Pr\{(\boldsymbol{X}, \boldsymbol{X'}) \in \mathcal{T}(Q_{XX'})\} \tag{C.5}$$

$$= M \cdot \frac{|\mathcal{T}(Q_{XX'})|}{|\mathcal{T}(Q_X)|^2} \tag{C.6}$$

$$\doteq M \cdot \frac{\exp\{nH_Q(X,X')\}}{e^{2nH_Q(X)}} \tag{C.7}$$

$$= \exp\{n[R - I_Q(X;X')]\}, \tag{C.8}$$

where the unconditional expectation in the second and third lines corresponds to uniform random selection across the whole class of fixed composition codes, not just to $\mathcal{G}_\epsilon$. It follows then that in the ensemble of all randomly selected codes within $\mathcal{G}_\epsilon$:

$$\begin{aligned}
&\Pr \bigcup_{Q_{XX'}} \left\{ \mathcal{C}_n : \ \frac{1}{M} \sum_{m=0}^{M-1} N_m(Q_{XX'}, \mathcal{C}_n) > \exp\{n[R - I_Q(X;X') + \epsilon/2]\} \middle| \mathcal{C}_n \in \mathcal{G}_\epsilon \right\} \\
\leq \ & \sum_{Q_{XX'}} \Pr \left\{ \mathcal{C}_n : \ \frac{1}{M} \sum_{m=0}^{M-1} N_m(Q_{XX'}, \mathcal{C}_n) > \exp\{n[R - I_Q(X;X') + \epsilon/2]\} \middle| \mathcal{C}_n \in \mathcal{G}_\epsilon \right\} \\
\doteq \ & \sum_{Q_{XX'}} \frac{\overline{N(Q_{XX'})}}{\exp\{n[R - I_Q(X;X') + \epsilon/2]\}} \\
\doteq \ & \sum_{Q_{XX'}} e^{-n\epsilon/2} \\
\leq \ & (n+1)^{|\mathcal{X}|^2} \cdot e^{-n\epsilon/2} \to 0,
\end{aligned} \tag{C.9}$$

which means that there exists a code in $\mathcal{G}_\epsilon$ (and in fact, for almost every such code),

$$\frac{1}{M} \sum_{m=0}^{M-1} N_m(Q_{XX'}, \mathcal{C}_n) \leq \exp\{n[R - I_Q(X;X') + \epsilon/2]\} \quad \forall Q_{XX'}. \tag{C.10}$$

For a given such code and every given $Q_{XX'}$, there must then exist at least $(1 - e^{-n\epsilon/2}) \cdot M$ values of m such that

$$N_m(Q_{XX'}, \mathcal{C}_n) \leq \exp\{n[R - I_Q(X;X') + \epsilon]\}. \tag{C.11}$$

Upon eliminating the exceptional codewords from the code, for all $Q_{XX'}$, one ends up with at least $[1 - (n+1)^{|\mathcal{X}|^2} e^{-n\epsilon/2}] \cdot M$ for which

$$N_m(Q_{XX'}, \mathcal{C}_n) \leq \exp\{n[R - I_Q(X;X') + \epsilon]\} \quad \forall Q_{XX'}. \tag{C.12}$$

Let $\mathcal{C}_n'$ denote the sub–code formed by these $[1 - (n+1)^{|\mathcal{X}|^2} e^{-n\epsilon/2}] \cdot M$ remaining codewords. Since $N_m(Q_{XX'}, \mathcal{C}_n') \leq N_m(Q_{XX'}, \mathcal{C}_n)$, then the sub–code $\mathcal{C}_n'$ certainly satisfies

$$N_m(Q_{XX'}, \mathcal{C}_n') \leq \exp\{n[R - I_Q(X;X') + \epsilon]\} \quad \forall m, Q_{XX'}. \tag{C.13}$$

Finally, observe that since $N_m(Q_{XX'}, \mathcal{C}_n')$ is a non–negative integer, then for $Q_{XX'}$ with $R - I_Q(X;X') + \epsilon < 0$, the last inequality means $N_m(Q_{XX'}, \mathcal{C}_n') = 0$, in which case the r.h.s. of the last equation becomes $N^*(Q_{XX'})$. Thus, we have shown that there exists a code $\mathcal{C}_n'$ of size $M' = [1 - (n+1)^{|\mathcal{X}|^2} e^{-n\epsilon/2}] \cdot e^{nR}$ for which all codewords satisfy $N_m(Q_{XX'}, \mathcal{C}_n') \leq N^*(Q_{XX'})$ for all joint types $Q_{XX'}$.

References

[1] I. Csiszár, "Joint source–channel error exponent," *Problems of Control and Information Theory*, vol. 9, no. 5, pp. 315–328, 1980.

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

[3] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of a discrete memoryless channel," *Proc. ISIT '77*, p. 107 (abstract), Cornell University, Itacha, New York, U.S.A., 1977.

[4] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 35–43, January 1995.

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

[6] J. Hui, *Fundamental issues of multiple accessing*, Ph.D. dissertation, MIT, 1983.

[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, "Relations between random coding exponents and the statistical physics of random codes," *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 83–92, January 2009.

[9] N. Merhav, "Erasure/list exponents for Slepian–Wolf decoding," *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4463–4471, August 2014.

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

[11] N. Merhav, "Universal decoding for source–channel coding with side information," submitted to *IEEE Trans. Inform. Theory*, July 2015. Available on–line at `http://arxiv.org/pdf/1505.01255.pdf`

[12] N. Merhav, "List decoding – random coding exponents and expurgated exponents," *IEEE Trans. Inform. Theory*, vol. 60, no. 11, pp. 6749–6759, November 2014.

[13] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1953-1967, November 1994.

[14] P. Ruján, "Finite temperature error–correcting codes," *Phys. Rev. Let.*, vol. 70, no. 19, pp. 2968–2971, May 1993.

[15] 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.

[16] E. C. Song, P. Cuff and H. V. Poor, "The likelihood encoder for lossy compression," [`http://arxiv.org/abs/1408.4522`], 2014.

[17] M. H. Yassaee, M. R. Aref and A. Gohari, "A technique for deriving one–shot achievability results in network information theory," [http://arxiv.org/abs/1303.0696], 2013.