IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Erasure/List Exponents for Slepian-Wolf Decoding

Neri Merhav

CCIT Report #830
May 2013

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

CCIT Report #830 May 2013

Erasure/List Exponents for Slepian–Wolf Decoding*

Neri Merhav

May 24, 2013

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

Abstract

We analyze random coding error exponents associated with erasure/list Slepian–Wolf decoding using two different methods and then compare the resulting bounds. The first method follows the well known techniques of Gallager and Forney and the second method is based on a technique of distance enumeration, or more generally, type class enumeration, which is rooted in the statistical mechanics of a disordered system that is related to the random energy model (REM). The second method is guaranteed to yield exponent functions which are at least as tight as those of the first method, and it is demonstrated that for certain combinations of coding rates and thresholds, the bounds of the second method are strictly tighter than those of the first method, by an arbitrarily large factor. In fact, the second method may even yield an infinite exponent at regions where the first method gives finite values. We also discuss the option of variable–rate Slepian–Wolf encoding and demonstrate how it can improve on the resulting exponents.

Index Terms: Slepian–Wolf coding, error exponents, erasure/list decoding, phase transitions.

*This research was supported by the Israeli Science Foundation (ISF) grant no. 412/12.

1 Introduction

The celebrated paper by Slepian and Wolf [14] has ignited a long–lasting, intensive research activity on separate source coding and joint decoding of correlated sources, during the last four decades. Besides its extensions in many directions, some of the more recent studies have been devoted to further refinements of performance analysis, such as exponential bounds on the decoding error probability. In particular, Gallager [9] derived a lower bound on the achievable random coding error exponent pertaining to random binning (henceforth, random binning exponent), using a technique that is very similar to that of his derivation of the ordinary random coding error exponent [8, Sections 5.5–5.6]. This random binning exponent was later shown by Csiszár, Körner and Marton [2], [4] to be universally achievable. The work of Csiszár and Körner [3] is about a universally achievable error exponent using linear codes as well as a non–universal, expurgated exponent which is improved at high rates. More recently, Csiszár [1] and Oohama and Han [13] have derived error exponents for the more general setting of coded side information. For large rates at one of the encoders, Kelly and Wagner [10] improved upon these results, but they did not consider the general case.

Since Slepian–Wolf decoding is essentially an instance of channel decoding, we find it natural to examine its performance also in the framework of generalized channel decoders, that is, decoders with an erasure/list option. Accordingly, this paper is about the analysis of random binning exponents associated with generalized decoders. It should be pointed out that error exponents for list decoders of the Slepian–Wolf encoders were already analyzed in [5], but in that work, it was assumed that the list size is fixed (independent of the block length) and deterministic. In this paper, on the hand, we analyze achievable trade-offs between random binning exponents associated with erasure/list decoders in the framework similar to that of Forney [7]. This means, among other things, that the erasure and list options are treated jointly, on the same footing, using an optimum decision rule of a common form, and that in the list option, the list size is a random variable whose typical value might be exponentially large in the block length. The erasure option allows the decoder not to decode when the confidence level is not satisfactory. It can be motivated, for example, by the possibility of generating a rate-less Slepian–Wolf code (see also [6]), provided that there is at least some minimum amount of feedback.

We analyze random binning error exponents associated with erasure/list Slepian–Wolf decoding using two different methods and then compare the resulting bounds. The first method follows the well known techniques of Gallager [8] and Forney [7], whereas the second method is based on a technique of distance enumeration, or more generally, on type class enumeration. This method has already been used in previous work (see [11, Chapters 6–7] and references therein) and proved useful in obtaining bounds on error exponents which are always at least as tight[1] (and in many cases, strictly tighter) than those obtained in the traditional methods of the information theory literature. This technique is rooted in the statistical mechanics of certain models of disordered magnetic materials. While in the case of ordinary random coding, the parallel statistical–mechanical model is the random energy model (REM) [12, Chapters 5–6], [11, Chapters 6–7], here, since random binning is considered, the parallel statistical–mechanical model is slightly different, but related. We will refer to this model as the *random dilution model* (RDM) for reasons that will become apparent in the sequel.

As mentioned in the previous paragraph, the type class enumeration method is guaranteed to yield an exponent function which is at least as tight as that of the classical method. But it is also demonstrated that for certain combinations of coding rates and thresholds of the erasure/list decoder, the exponent of the type class enumeration method is strictly tighter than that of the ordinary method. In fact, the gap between them (i.e., their ratio) can be arbitrarily large, and even strictly infinite. In other words, for a small enough threshold (pertaining to list decoding), the former exponent can be infinite while the latter is finite.

While the above described study is carried out for fixed–rate Slepian–Wolf encoding, we also demonstrate how variable–rate encoding (with a certain structure) can strictly improve on the random binning exponents. This is shown in the context of the exponents derived using the Forney/Gallager method, but a similar generalization can be carried out using the other method.

The outline of the paper is as follows. In Section 2, we provide notation conventions and define the objectives of the paper more formally. In Section 3, we derive the random binning exponents using the Forney/Gallager method, and in Section 4, we extend this analysis to allow variable

[1]It should be pointed out that in [15], another version of the type class enumeration method, which is guaranteed to yield the exact random coding exponents, was developed. This method, however, is much more difficult to implement and it gives very complicated expressions.

rate coding. Finally, in Section 5, after a short background on the relevant statistical–mechanical model (Subsection 5.1), we use the type class enumeration technique, first in the binary case (Subsection 5.2), then compare the resulting exponents to those of Section 3 (Subsection 5.3), and finally, generalize the analysis to a general pair of correlated finite alphabet memoryless sources (Subsection 5.4).

2 Notation Conventions, Problem Formulation and Background

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.

For a given vector $\boldsymbol{x}$, let $\hat{P}_{\boldsymbol{x}}$ denote the empirical distribution, that is, the vector $\{\hat{P}_{\boldsymbol{x}}(x),\ x \in \mathcal{X}\}$, where $\hat{P}_{\boldsymbol{x}}(x)$ is the relative frequency of the letter x in the vector $\boldsymbol{x}$. Let $\mathcal{T}(\boldsymbol{x})$ denote its type class of $\boldsymbol{x}$, namely, the set $\{\boldsymbol{x}' :\ \hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}\}$. The empirical entropy associated with $\boldsymbol{x}$, denoted $\hat{H}_{\boldsymbol{x}}(X)$, is the entropy associated with the empirical distribution $\hat{P}_{\boldsymbol{x}}$. Similarly, for a pair of vectors $(\boldsymbol{x}, \boldsymbol{y})$, the empirical joint distribution $\hat{P}_{\boldsymbol{xy}}$ is the matrix $\{\hat{P}_{\boldsymbol{xy}}(x, y),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$ of relative frequencies of symbol pairs $\{(x, y)\}$. The conditional type class $\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})$ is the set $\{\boldsymbol{x}' :\ \hat{P}_{\boldsymbol{x}'\boldsymbol{y}} = \hat{P}_{\boldsymbol{xy}}\}$. The empirical conditional entropy of $\boldsymbol{x}$ given $\boldsymbol{y}$, denoted $\hat{H}_{\boldsymbol{xy}}(X|Y)$, is the conditional entropy of X given Y, associated with the joint empirical distribution $\{\hat{P}_{\boldsymbol{xy}}(x, y)\}$.

The expectation operator will be denoted by $\boldsymbol{E}\{\cdot\}$. Logarithms and exponents will be understood to be taken to the natural base unless specified otherwise. The indicator function will be denoted by $\mathcal{I}(\cdot)$. The notation function $[t]_+$ will be defined as $\max\{t, 0\}$. For two positive sequences, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ will mean asymptotic equivalence in the exponential scale, that is, $\lim_{n\to\infty} \frac{1}{n} \log(\frac{a_n}{b_n}) = 0$. Similarly, $a_n \stackrel{\cdot}{\leq} b_n$ will mean $\limsup_{n\to\infty} \frac{1}{n} \log(\frac{a_n}{b_n}) \leq 0$, and so on.

2.2 Problem Formulation and Background

Let $\{(X_i, Y_i)\}_{i=1}^n$ be n independent copies of a random vector (X, Y), distributed according to a given probability mass function $P(x, y)$, where x and y take on values in finite alphabets, $\mathcal{X}$ and $\mathcal{Y}$, respectively. The source vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, which is a generic realization of $\boldsymbol{X} = (X_1, \ldots, X_n)$, is compressed at the encoder by random binning, that is, each n–tuple $\boldsymbol{x} \in \mathcal{X}^n$ is randomly and independently assigned to one out of $M = e^{nR}$ bins, where R is the coding rate in nats per symbol. Given a realization of the random partitioning into bins (revealed to both the encoder and the decoder), let $f : \mathcal{X}^n \to \{0, 1, \ldots, M-1\}$ denote the encoding function, i.e., $z = f(\boldsymbol{x})$ is the encoder output. Accordingly, the inverse image of z, defined as $f^{-1}(z) = \{\boldsymbol{x} : \ f(\boldsymbol{x}) = z\}$, is the bin of all source vectors mapped by the encoder into z. The decoder has access to z and to $\boldsymbol{y} = (y_1, \ldots, y_n)$, which is a realization of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, namely, the side information at the decoder.

Following [7], we consider a decoder with an erasure/list option, defined as follows. Let $P(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^n P(x_i, y_i)$ denote the probability of the event $\{\boldsymbol{X} = \boldsymbol{x}, \ \boldsymbol{Y} = \boldsymbol{y}\}$ and let T be a given real valued parameter. The decoding rule is as follows. For every $\hat{\boldsymbol{x}} \in f^{-1}(z)$, if

$$\frac{P(\hat{\boldsymbol{x}}, \boldsymbol{y})}{\sum_{\boldsymbol{x}' \in f^{-1}(z) \setminus \{\hat{\boldsymbol{x}}\}} P(\boldsymbol{x}', \boldsymbol{y})} \geq e^{nT}, \tag{1}$$

then $\hat{\boldsymbol{x}}$ is referred to as a *candidate*. If there are no candidates, an erasure is declared, namely, the decoder acts in its erasure mode. If there is exactly one candidate, $\hat{\boldsymbol{x}}$, then this is the estimate that the decoder produces, just like in ordinary decoding. Finally, if there is more than one candidate, then the decoder operates in the list mode and it outputs the list of all candidates. Obviously, for $T \geq 0$, the list can contain at most one candidate. The list may contain two candidates or more only for sufficiently small negative values of T.

Forney [7] used the Neymann–Pearson lemma, in an analogous channel coding setting, to show that the above rule simultaneously gives rise to: (i) an optimum trade-off between the probability of erasure and the probability of decoding error, in the erasure mode, and (ii) an optimum trade-off between the probability of list error and the expected number of incorrect candidates on the list, in the list mode. Our goal, in this paper, is to assess the exponential rates associated with these trade-offs.

3 Error Exponent Analysis Based on the Gallager/Forney Method

Similarly as in [7], we define the event $\mathcal{E}_1$ as the event that the correct source vector $\boldsymbol{x}$ is not a candidate, that is,

$$\frac{P(\boldsymbol{x},\boldsymbol{y})}{\sum_{\boldsymbol{x}'\in f^{-1}(z)\setminus\{\boldsymbol{x}\}} P(\boldsymbol{x}',\boldsymbol{y})} < e^{nT}. \tag{2}$$

We next derive a lower bound on the exponential rate $E_1(R,T)$ of the average probability of $\mathcal{E}_1$, where the averaging is with respect to (w.r.t.) the ensemble of random binnings. The other exponent, $E_2(R,T)$ (of decoding error in the erasure option, or the expected list size in list option) will then be given by $E_2(R,T) = E_1(R,T)+T$, similarly as in [7]. We now have the following chain of inequalities for any $s \geq 0$:

$$\begin{aligned}
\Pr\{\mathcal{E}_1\} &= \sum_{\boldsymbol{x},\boldsymbol{y}} P(\boldsymbol{x},\boldsymbol{y})\mathcal{I}\left\{\frac{e^{nT}\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]}{P(\boldsymbol{x},\boldsymbol{y})} > 1\right\} \\
&\leq \sum_{\boldsymbol{x},\boldsymbol{y}} P(\boldsymbol{x},\boldsymbol{y})\left[\frac{e^{nT}\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]}{P(\boldsymbol{x},\boldsymbol{y})}\right]^s \\
&= e^{nsT}\sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y})\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\right]^s. \qquad (3)
\end{aligned}$$

Now, let $\rho \geq s$ be another parameter. Then,

$$\begin{aligned}
\Pr\{\mathcal{E}_1\} &\leq e^{nsT}\sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y})\left(\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\right]^{s/\rho}\right)^{\rho} \qquad (4) \\
&\leq e^{nsT}\sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y})\left(\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{s/\rho}(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\right)^{\rho}. \qquad (5)
\end{aligned}$$

where we have used the inequality $(\sum_i a_i)^t \leq \sum_i a_i^t$ for $t \in [0,1]$. Taking now the expectation w.r.t. the randomness of the binning, and assuming that $\rho \leq 1$, we get

$$\begin{aligned}
\overline{\Pr\{\mathcal{E}_1\}} &\leq e^{nsT} \sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y}) \boldsymbol{E}\left\{\left(\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{s/\rho}(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\right)^{\rho}\right\} && (6)\\
&\leq e^{nsT} \sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y}) \left(\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{s/\rho}(\boldsymbol{x}',\boldsymbol{y})\boldsymbol{E}\{\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\}\right)^{\rho} && (7)\\
&= e^{nsT} \sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y}) \left(\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{s/\rho}(\boldsymbol{x}',\boldsymbol{y})e^{-nR}\right)^{\rho} && (8)\\
&= e^{-n(\rho R - sT)} \sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y}) \left(\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{s/\rho}(\boldsymbol{x}',\boldsymbol{y})\right)^{\rho} && (9)\\
&= e^{-n(\rho R - sT)} \sum_{\boldsymbol{y}} P(\boldsymbol{y}) \sum_{\boldsymbol{x}} P^{1-s}(\boldsymbol{x}|\boldsymbol{y}) \left(\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{s/\rho}(\boldsymbol{x}'|\boldsymbol{y})\right)^{\rho} && (10)\\
&\leq e^{-n(\rho R - sT)} \left[\sum_{y\in\mathcal{Y}} P(y) \sum_{x\in\mathcal{X}} P^{1-s}(x|y) \left(\sum_{x'\in\mathcal{X}} P^{s/\rho}(x'|y)\right)^{\rho}\right]^{n}. && (11)
\end{aligned}$$

Thus, after optimization over ρ and s, subject to the constraints $0 \leq s \leq \rho \leq 1$, we obtain

$$\overline{\Pr\{\mathcal{E}\}} \leq e^{-nE_1(R,T)} \tag{12}$$

where

$$E_1(R,T) = \sup_{0\leq s\leq\rho\leq 1} [E_0(\rho,s) + \rho R - sT] \tag{13}$$

with

$$E_0(\rho,s) = -\ln\left[\sum_{y\in\mathcal{Y}} P(y) \sum_{x\in\mathcal{X}} P^{1-s}(x|y) \left(\sum_{x'\in\mathcal{X}} P^{s/\rho}(x'|y)\right)^{\rho}\right]. \tag{14}$$

A few elementary properties of the function $E_1(R,T)$ are the following.

1. $E_1(R,T)$ is jointly convex in both arguments. This follows directly from the fact that it is given by the supremum over a family of affine functions in (R,T). Clearly, $E_1(R,T)$ is increasing in R and decreasing in T.

2. At $T=0$, the optimum s is $\rho/(1+\rho)$, similarly as in [7] and [9]. Thus, as observed in [7], here too, the case $T=0$ is essentially equivalent (in terms of error exponents) to ordinary decoding, although operationally, there still might be erasures in this case.

3. For a given T, the infimum of R such that $E_1(R,T) > 0$ is

$$R_{\min}(T) = \inf_{0 \le s \le \rho \le 1} \frac{sT - E_0(\rho, s)}{\rho}, \tag{15}$$

which is a concave increasing function. At $T = 0$,

$$R_{\min}(0) = -\sup_{0 \le \rho \le 1} \frac{E_0\left(\rho, \frac{\rho}{1+\rho}\right)}{\rho} = -\lim_{\rho \to 0} \frac{E_0\left(\rho, \frac{\rho}{1+\rho}\right)}{\rho} = -\frac{\partial}{\partial \rho} E_0\left(\rho, \frac{\rho}{1+\rho}\right)\Big|_{\rho=0} = H(X|Y).$$

4. For a given R, the supremum of T such that $E_1(R,T) > 0$ is

$$T_{\max}(R) = \sup_{0 \le s \le \rho \le 1} \frac{\rho R + E_0(\rho, s)}{s}, \tag{16}$$

which is a convex increasing function, the inverse of $R_{\min}(T)$.

Additional properties can be found similarly as in [7], but we will not delve into them here.

4 Extension to Variable–Rate Slepian–Wolf Coding

A possible extension of the above error exponent analysis allows variable rate coding. In this section, we demonstrate how the flexibility of variable–rate coding can improve the error exponents.

Consider an encoder that first sends a relatively short header that encodes the type class of $\boldsymbol{x}$ (using a logarithmic number of bits), and then a description of $\boldsymbol{x}$ within its type class, using a random bin $z = f(\boldsymbol{x})$ in the range $\{0, 1, \ldots, \exp[nR(\boldsymbol{x})] - 1\}$, where $R(\boldsymbol{x}) > 0$ depends on $\boldsymbol{x}$ only via the type class of $\boldsymbol{x}$. The bin z for every $\boldsymbol{x}$ in its type class is selected independently at random with a uniform probability distribution $P(z) = e^{-nR(\boldsymbol{x})}$. The average coding rate would be, of course, $R = \boldsymbol{E}\{R(\boldsymbol{X})\}$ (neglecting the rate of the header). For example, consider an additive rate function[2] $R(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} r(x_i)$. Thus, $R = \boldsymbol{E}\{r(X)\} = \sum_{x \in \mathcal{X}} P(x) r(x)$. Extending the above error exponent analysis, one readily obtains[3]

$$\tilde{E}_1(R,T) = \sup_{0 \le s \le \rho \le 1} \sup_{\{\boldsymbol{r}:\ \boldsymbol{E}\{r(X)\} \le R,\ r(x) > 0\ \forall\ x \in \mathcal{X}\}} [\tilde{E}_0(\rho, s) - sT], \tag{17}$$

[2]The reason for choosing a rate function with this simple structure is that it allows to easily generalize the analysis in the Gallager/Forney style and obtain single–letter expressions without recourse to the method of types. More general rate functions, that depend on the type class of $\boldsymbol{x}$ in an arbitrary manner, are still manageable, but require the method of types.

[3]Observe that here $\Pr\{f(\boldsymbol{x}') = f(\boldsymbol{x})\} = e^{-nR(\boldsymbol{x}')}$ whenever $e^{nR(\boldsymbol{x}')} < f(\boldsymbol{x})$ and $\Pr\{f(\boldsymbol{x}') = f(\boldsymbol{x})\} = 0$ elsewhere, thus $\Pr\{f(\boldsymbol{x}') = f(\boldsymbol{x})\} \le e^{-nR(\boldsymbol{x}')}$ everywhere.

where $\boldsymbol{r} \stackrel{\Delta}{=} \{r(x),\ x \in \mathcal{X}\}$ and where $\tilde{E}_0(\rho, s)$ is defined as

$$\tilde{E}_0(\rho, s) = -\ln \left[\sum_{y \in \mathcal{Y}} P(y) \sum_{x \in \mathcal{X}} P^{1-s}(x|y) \left(\sum_{x' \in \mathcal{X}} P^{s/\rho}(x'|y) e^{-r(x')} \right)^{\rho} \right]. \tag{18}$$

It is interesting to find the optimum rate assignment $\boldsymbol{r} = \{r(x)\ x \in \mathcal{X}\}$ that maximizes the exponent. Consider, for example, the case where R and T are such that $E_1(R, T)$ is achieved by $\rho = 1$. Then,

$$\begin{aligned} e^{-E_0(1,s)} &= \sum_{y \in \mathcal{Y}} P(y) \sum_{x \in \mathcal{X}} P^{1-s}(x|y) \sum_{x' \in \mathcal{X}} P^{s}(x'|y) e^{-r(x')} && (19) \\ &= \sum_{x \in \mathcal{X}} F(x) e^{-r(x)} && (20) \end{aligned}$$

where

$$F(x) \stackrel{\Delta}{=} \sum_{y \in \mathcal{Y}} P(y) P^s(x|y) \sum_{x' \in \mathcal{X}} P^{1-s}(x'|y). \tag{21}$$

Our task now is to minimize $\sum_{x \in \mathcal{X}} F(x) e^{-r(x)}$ subject to the constraints $\sum_{x \in \mathcal{X}} P(x) r(x) \leq R$ and $r(x) > 0$ for all $x \in \mathcal{X}$, which is a standard convex program. For simplicity, let us first ignore the constraints $r(x) > 0$, $x \in \mathcal{X}$, and assume that the parameters of the problem are such that the resulting solution will satisfy these positivity constraints anyway. Then,

$$r^*(x) = \lambda + \ln \frac{F(x)}{P(x)}, \tag{22}$$

where λ is determined by the average rate constraint, that is

$$\begin{aligned} \lambda &= R + \sum_{x \in \mathcal{X}} P(x) \ln \frac{P(x)}{F(x)} && (23) \\ &= R + D(P\|Q) - \ln \left[\sum_{x \in \mathcal{X}} F(x) \right], && (24) \end{aligned}$$

where

$$Q(x) = \frac{F(x)}{\sum_{x' \in \mathcal{X}} F(x')}. \tag{25}$$

Thus,

$$r^*(x) = R + D(P\|Q) + \ln \frac{Q(x)}{P(x)}. \tag{26}$$

We see that fixed–rate coding is optimum only if $P(x)$ happens to be proportional to $F(x)$, namely, $P = Q$ (which is the case, for example, when $s = 1$). Upon substituting $\boldsymbol{r}^* = \{r^*(x),\ x \in \mathcal{X}\}$ back

into the objective function, we obtain

$$e^{-\tilde{E}_0(1,s)} = \sum_{x\in\mathcal{X}} F(x)\exp\{-\lambda - \ln[F(x)/P(x)]\} \quad (27)$$

$$= \sum_{x\in\mathcal{X}} P(x)e^{-\lambda} = e^{-\lambda}, \quad (28)$$

and so,

$$\tilde{E}_0(1,s) = \lambda \quad (29)$$

$$= R + D(P\|Q) - \ln\left[\sum_{x\in\mathcal{X}} F(x)\right] \quad (30)$$

$$= R + D(P\|Q) - \ln\left[\sum_{y\in\mathcal{Y}} P(y)\sum_{x\in\mathcal{X}} P^{1-s}(x|y)\sum_{x'\in\mathcal{X}} P^s(x'|y)\right] \quad (31)$$

$$= E_0(1,s) + D(P\|Q). \quad (32)$$

The term $D(P\|Q)$ then represents the improvement we have obtained upon passing from fixed–rate coding to variable–rate coding with an additive rate function. This is true for a given s. However, after re–optimizing the bound over s, the improvement can be even larger. When $R + D(P\|Q) + \ln[Q(x)/P(x)]$ are not all positive, the optimum solution is given by

$$r^*(x) = \left[\ln\frac{Q(x)}{P(x)} + \mu\right]_+ \quad (33)$$

where μ is the (unique) solution to the equation

$$\sum_{x\in\mathcal{X}} P(x)\left[\ln\frac{Q(x)}{P(x)} + \mu\right]_+ = R. \quad (34)$$

For $\rho < 1$, the optimization over $\boldsymbol{r}$ is less trivial, but it can still be carried out at least numerically.

5 Error Exponent Analysis Using Type Class Enumeration

5.1 A Brief Background in Statistical Mechanics

This subsection can be skipped without essential loss of continuity, however, we believe that before getting into the detailed technical derivation, it would be instructive to give a brief review of the statistical–mechanical models that are at the basis of the type class enumeration method.

In ordinary random coding (as opposed to random binning), the derivations of bounds on the error probability (especially in the methods of Gallager and Forney) are frequently associated with

expressions of the form $\sum_{\boldsymbol{x}\in\mathcal{C}} P^{\beta}(\boldsymbol{y}|\boldsymbol{x})$, where $\mathcal{C}$ is (randomly selected) codebook and $\beta > 0$ is some parameter. As explained in [11, Chap. 6], this can be viewed, from the statistical–mechanical perspective, as a partition function

$$Z(\beta) = \sum_{\boldsymbol{x}\in\mathcal{C}} e^{-\beta E(\boldsymbol{x},\boldsymbol{y})}, \tag{35}$$

where β plays the role of inverse temperature and where the energy function (Hamiltonian) is $E(\boldsymbol{x},\boldsymbol{y}) = -\ln P(\boldsymbol{y}|\boldsymbol{x})$. Since the codewords are selected independently at random, then for a given $\boldsymbol{y}$, the energies $\{E(\boldsymbol{x},\boldsymbol{y}),\ \boldsymbol{x}\in\mathcal{C}\}$ are i.i.d. random variables. This is, in principle, nothing but the *random energy model* (REM), a well known model in statistical mechanics of disordered magnetic materials (spin glasses), which exhibits a phase transition: below a certain critical temperature ($\beta > \beta_c$), the system freezes in the sense that the partition function is exponentially dominated by a subexponential number of configurations at the ground–state energy (zero thermodynamical entropy). This phase is called the *frozen phase* or the *glassy phase*. The other phase, $\beta < \beta_c$, is called the *paramagnetic phase* (see more details in [12, Chap. 5]). Accordingly, the resulting exponential error bounds associated with random coding 'inherit' this phase transition (see [11] and references therein).

In random binning the situation is somewhat different. As we have seen in Section 3, here the bound involves an expression like $\sum_{\boldsymbol{x}'} P^{\beta}(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]$. The source vectors $\{\boldsymbol{x}'\}$ that participate in the summation are now deterministic, but the random ingredient is the function f. The analogous statistical–mechanical model is then encoded into the partition function

$$Z(\beta) = \sum_{\boldsymbol{x}} I(\boldsymbol{x})\cdot e^{-\beta E(\boldsymbol{x},\boldsymbol{y})}, \tag{36}$$

where $\{I(\boldsymbol{x}),\ \boldsymbol{x}\in\mathcal{X}^n\}$ are i.i.d. binary random variables, taking on values in $\{0,1\}$, where $\Pr\{I(\boldsymbol{x}) = 1\} = e^{-nR}$. In other words, $Z(\beta)$ is a randomly diluted version of the full partition function $\sum_{\boldsymbol{x}} e^{-\beta E(\boldsymbol{x},\boldsymbol{y})}$, where each configuration $\boldsymbol{x}$ 'survives' with probability e^{-nR} or is discarded with probability $1-e^{-nR}$. Accordingly, we refer to this model as the *random dilution model* (RDM). To the best of our knowledge, such a model has not been used in statistical mechanics thus far, but it can be analyzed in the very same fashion, and it is easy to see that it also exhibits a glassy phase transition (depending on R). In fact, the RDM can be considered as a variant of the REM, where the configurational energies are $E(\boldsymbol{x},\boldsymbol{y}) + \phi(\boldsymbol{x})$, where $\phi(\boldsymbol{x}) = 0$ with probability e^{-nR} and

$\phi(\boldsymbol{x}) = \infty$ with probability $1-e^{-nR}$. Thus, $\phi(\boldsymbol{x})$ can be thought of as disordered potential function, associated with long–range interactions, with infinite spikes that forbid access to certain points in the configuration space.

5.2 The Binary Case

Let us return to the fixed–rate regime. It is instructive to begin from the relatively simple special case where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are correlated binary symmetric sources (BSS's), that is,

$$P(x,y) = \begin{cases} (1-p)/2 & x = y \\ p/2 & x \neq y \end{cases} \qquad x, y \in \{0,1\} \tag{37}$$

We begin similarly as in Section 3: Our starting point is the same bound as in the last line of eq. (3), specialized to the binary case considered here, where we also take the ensemble average:

$$\begin{aligned}
\overline{\Pr\{\mathcal{E}_1\}} &\leq e^{nsT} \sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y}) \cdot \boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]\right]^s\right\} && (38)\\
&= e^{nsT} \sum_{\boldsymbol{y}} P(\boldsymbol{y}) \left[\sum_{\boldsymbol{x}} P^{1-s}(\boldsymbol{x}|\boldsymbol{y})\right] \cdot \boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}'|\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]\right]^s\right\} && (39)\\
&= e^{nsT} \sum_{\boldsymbol{y}} 2^{-n} \left[p^{1-s} + (1-p)^{1-s}\right]^n \cdot \boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}'|\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]\right]^s\right\} && (40)\\
&= e^{nsT} \left[p^{1-s} + (1-p)^{1-s}\right]^n \cdot \boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}'|\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]\right]^s\right\} && (41)
\end{aligned}$$

where the last step is justified by the fact that the expectation term is independent of $\boldsymbol{y}$, as will be seen shortly. Now,

$$\begin{aligned}
\boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}'|\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]\right]^s\right\} &\doteq \sum_{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})} P^s(\boldsymbol{x}'|\boldsymbol{y})\boldsymbol{E}\{N^s(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y})\} && (42)\\
&= (1-p)^{ns} \sum_{\delta} \left(\frac{p}{1-p}\right)^{ns\delta} \boldsymbol{E}\{N^s(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y})\} && (43)
\end{aligned}$$

where δ is the normalized Hamming distance, the summation is over the set $\{0, 1/n, 2/n, \ldots, 1-1/n, 1\}$, and $N(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y}) = \sum_{\boldsymbol{x}'\in\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})} \mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]$. Now, $N(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y})$ is the sum of $|\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})| \doteq \exp\{n\hat{H}_{\boldsymbol{x}\boldsymbol{y}}(X|Y)\}$ i.i.d. binary random variables $\{\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]\}$ with $\Pr\{f(\boldsymbol{x}') = f(\boldsymbol{x})\} = e^{-nR}$.

Thus, similarly as in [11, Sect. 6.3]

$$\begin{aligned} \boldsymbol{E}\{N^s(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y})\} &\doteq \begin{cases} \exp\{ns[h(\delta)-R]\} & h(\delta)\geq R \\ \exp\{n[h(\delta)-R]\} & h(\delta)<R \end{cases} && (44)\\ &= \exp\{n(s[h(\delta)-R]-(1-s)[R-h(\delta)]_+)\} && (45) \end{aligned}$$

and so

$$\begin{aligned} \overline{\Pr\{\mathcal{E}_1\}} &\leq e^{nsT}\left[p^{1-s}+(1-p)^{1-s}\right]^n(1-p)^{ns}\sum_{\delta}\left(\frac{p}{1-p}\right)^{ns\delta}\times \\ &\quad \exp\{n(s[h(\delta)-R]-(1-s)[R-h(\delta)]_+)\} && (46)\\ &\doteq e^{nsT}\left[p^{1-s}+(1-p)^{1-s}\right]^n(1-p)^{ns}e^{-nL(R,s)} && (47) \end{aligned}$$

where $L(R,s)\stackrel{\Delta}{=}\min_{0\leq\delta\leq 1}L(R,s,\delta)$ with

$$L(R,s,\delta)\stackrel{\Delta}{=}s\delta\ln\frac{1-p}{p}+s[R-h(\delta)]+(1-s)[R-h(\delta)]_+. \tag{48}$$

Standard optimization of $L(R,s,\delta)$ gives the following result (see Appendix A for the details). Define the sets (see also Fig. 1)

$$\begin{aligned} A &= \{(s,R): 0\leq s\leq 1,\ R>h(p_s)\} && (49)\\ B &= \{(s,R): 0\leq s\leq 1,\ h(p)<R\leq h(p_s)\} && (50)\\ C &= \{(s,R): 0\leq s\leq 1,\ R\leq h(p)\} && (51)\\ D &= \{(s,R): s>1,\ R>h(p)\} && (52)\\ E &= \{(s,R): s>1,\ R(s)<R\leq h(p)\} && (53)\\ F &= \{(s,R): s>1,\ h(p_s)<R\leq R(s)\} && (54)\\ G &= \{(s,R): s>1,\ R\leq h(p_s)\}. && (55) \end{aligned}$$

Then,

$$L(R,s)=\begin{cases} s[p\ln\frac{1-p}{p}+R-h(p)] & (s,R)\in C\cup F\cup G \\ sh^{-1}(R)\ln\frac{1-p}{p} & (s,R)\in B \\ sp_s\ln\frac{1-p}{p}+R-h(p_s) & (s,R)\in A\cup D\cup E \end{cases} \tag{56}$$

Finally, the exponent of $\overline{\Pr\{E_1\}}$ is lower bounded by

$$E_1'(R,T)=\sup_{s\geq 0}\left\{L(R,s)+s\ln\frac{1}{1-p}-\ln[p^{1-s}+(1-p)^{1-s}]-sT\right\}. \tag{57}$$

Equivalently, $E_1'(R,T)$ can be presented as follows:

$$E_1'(R,T) = \sup_{s\geq 0} E_1'(R,T,s) \tag{58}$$

where

$$E_1'(R,T,s) = \begin{cases} s(R-T) - \ln[p^{1-s} + (1-p)^{1-s}] & (s,R) \in C \cup F \cup G \\ s[R - T + D(h^{-1}(R)\|p)] - \ln[p^{1-s} + (1-p)^{1-s}] & (s,R) \in B \\ R - sT - \ln[p^s + (1-p)^s] - \ln[p^{1-s} + (1-p)^{1-s}] & (s,R) \in A \cup D \cup E \end{cases} \tag{59}$$

Fig. 1 depicts a phase diagram of the function $L(R,s)$. This function inherits phase transitions associated with the analogous statistical–mechanical model – the RDM. The strip defined by $s \geq 0$ and $0 \leq R \leq \ln 2$ is divided into seven regions, labeled by the letters A–G as defined above. There are three main phases that are separated by solid lines, which differ in terms of the expression of $L(R,s)$. The phase $C \cup F \cup G$ is the phase where typical realiztions of the random binning ensemble dominate the partition function (that is, conditional type classes of size less than e^{nR} contain no matching bin, whereas conditional type classes of larger size have an exponentially typical number of bin matches), phase B is the glassy phase, and phase $A \cup D \cup E$ is the phase where the conditional small type classes dominate the partition function (unlike in phase $C \cup F \cup G$). A secondary partition into sub–phases (dashed lines) correspond to different shapes of the objective function $L(R,s,\delta)$. In regions A, B, C ($s \leq 1$), the derivative of the objective function has a positive jump at $\delta = h^{-1}(R)$, and the minimizer is smaller than $h^{-1}(R)$, equal to $h^{-1}(R)$, and larger than $h^{-1}(R)$, respectively. In regions D, E, F and G ($s > 1$), the derivative of $L(R,s,\delta)$ w.r.t. δ has a negative jump at $\delta = h^{-1}(R)$, In regions E and F, this jump is from a positive derivative to a negative derivative, meaning that $\delta = h^{-1}(R)$ is a (non–smooth) local maximum and there are two local minima, one at $\delta = p < h^{-1}(p)$ and one at $\delta = p_s > h^{-1}(R)$. In region E, the local minimum at $\delta = p_s$ is smaller than the local minimum at $\delta = p$ and in region F it is vice versa. In region G there is only one local minimum at $\delta = p$ and in region D there is only one local minimum at $\delta = p_s$.

5.3 Comparison of the Exponents

The expression of $E_1'(R,T)$ should be compared with $E_1(R,T)$ specialized to the double BSS considered in Subsection 5.2, i.e.,

$$E_1(R,T) = \sup_{0\leq s\leq\rho\leq 1} \left\{\rho R - \ln[p^{1-s} + (1-p)^{1-s}] - \rho \ln[p^{s/\rho} + (1-p)^{s/\rho}] - sT\right\}. \tag{60}$$

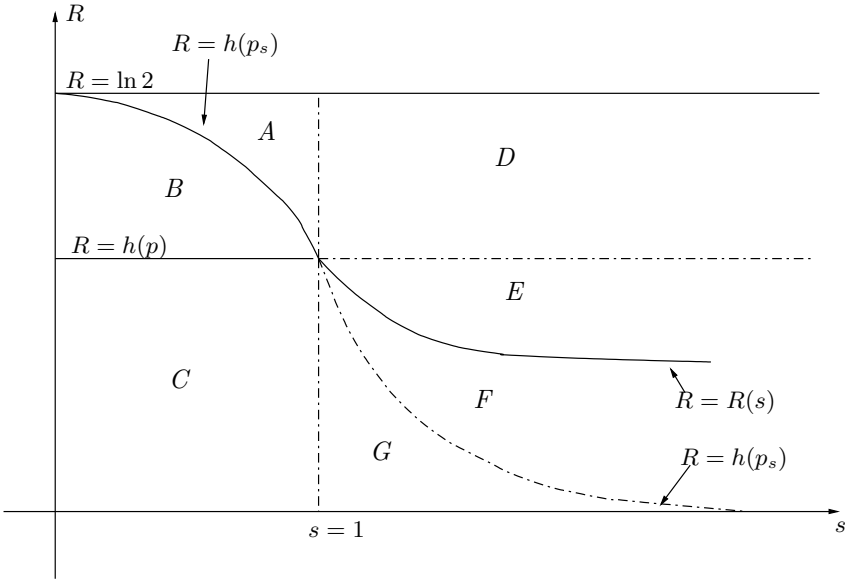

Figure 1: Phase diagram of the function $L(R, s)$.

Obviously, $E_1'(R,T) \geq E_1(R,T)$ since derivation of $E_1'(R,T)$ is guaranteed to be exponentially tight starting from (3), in contrast to the derivation of $E_1(R,T)$, which is associated with Jensen's inequality, as well as the inequality $(\sum_i a_i)^t \leq \sum_i a_i^t$, $0 \leq t \leq 1$, following [7].

To show an extreme situation of a strict inequality, $E_1'(R,T) > E_1(R,T)$, consider the case where $R > h(p)$ and $T < \ln[p/(1-p)] < 0$ (a list option). Then,

$$\begin{aligned}
E_1'(R,T) &\geq \lim_{s\to\infty}\left\{R - sT - \ln\left[(1-p)^s\left(1+\left[\frac{p}{1-p}\right]^s\right)\right] - \right. \\
&\quad \left. \ln\left[p^{1-s}\left(1+\left[\frac{1-p}{p}\right]^{1-s}\right)\right]\right\} && (61)\\
&= \lim_{s\to\infty}\left\{R - sT - s\ln(1-p) - \ln\left(1+\left[\frac{p}{1-p}\right]^s\right) - (1-s)\ln p - \right. \\
&\quad \left. \ln\left(1+\left[\frac{p}{1-p}\right]^{s-1}\right)\right\} && (62)\\
&= \lim_{s\to\infty}\{R - sT - s\ln(1-p) - (1-s)\ln p\} && (63)\\
&= \ln\frac{1}{p} + R + \lim_{s\to\infty} s\left[\ln\frac{p}{1-p} - T\right] && (64)\\
&= \infty. && (65)
\end{aligned}$$

On the other hand, in this case,

$$E_1(R,T) \leq R + |T| + 2 \max_{0\leq\alpha\leq 1} \{-\ln[p^\alpha + (1-p)^\alpha]\} \quad (66)$$

$$= R + |T| < \infty. \quad (67)$$

Another situation, where it is relatively easy to calculate the exponents is the limit of very weak correlation between the BSS's X and Y (in analogy to the notion of a very noisy channel [8, p. 147, Example 3]). Let $p = 1/2 - \epsilon$ for $|\epsilon| \ll 1$. In this case, a second order Taylor series expansion of the relevant functions (see Appendix B for the details) yields, for $h(p) \leq R \leq \ln 2$ and $T = -\tau\epsilon^2$, with $\tau > 4$ being fixed:

$$E_1(R,T) \leq (\tau + 2)\epsilon^2, \quad (68)$$

whereas

$$E_1'(R,T) \geq \left[\frac{\tau(\tau+8)}{16} - 1\right]\epsilon^2. \quad (69)$$

Now, observe that the upper bound on $E_1(R,T)$ is affine in τ, whereas the lower bound on $E_1'(R,T)$ is quadratic in τ, thus the ratio $E_1'(R,T)/E_1(R,T)$ can be made arbitrarily large for any sufficiently large $\tau > 4$.

In both examples, we took advantage of the fact that the range of optimization of s for $E_1'(R,T)$ includes all the positive reals, whereas for $E_1(R,T)$, it is limited to the interval $[0,1]$ due to the combination of using of Jensen's inequality (which requires $\rho \leq 1$) and the inequality $(\sum_i a_i)^t \leq \sum_i a_i^t$ (which requires $s \leq \rho$). Note that the second example is not a special case the first one, because in the first example, for $p = 1/2 - \epsilon$, $|T| > \ln[(1-p)/p] = O(\epsilon)$, whereas in the second example, $T = O(\epsilon^2)$.

5.4 Extension to General Finite Alphabet Memoryless Sources

In this subsection, we use the type class enumeration method for general finite alphabet sources S and Y. Consider the expression

$$\boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}') = f(\boldsymbol{x})]\right]^s\right\}$$

that appears upon taking the expectation over the last line of (3). Then, we have

$$\begin{aligned}
& \boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\right]^s\right\} && (70)\\
= \;& P^s(\boldsymbol{y})\boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}'|\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\right]^s\right\} && (71)\\
\leq \;& P^s(\boldsymbol{y})\sum_{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})} P^s(\boldsymbol{x}'|\boldsymbol{y})\boldsymbol{E}\left\{\left[\sum_{\tilde{\boldsymbol{x}}\in\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})} \mathcal{I}[f(\tilde{\boldsymbol{x}})=f(\boldsymbol{x})]\right]^s\right\} && (72)\\
\stackrel{\Delta}{=} \;& P^s(\boldsymbol{y})\sum_{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})} P^s(\boldsymbol{x}'|\boldsymbol{y})\boldsymbol{E}\left\{N^s(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y})\right\} && (73)
\end{aligned}$$

where $N(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y})$ is the (random) number of $\{\tilde{\boldsymbol{x}}\}$ in $\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})$ which belong to the same bin as $\boldsymbol{x}$. Now,

$$\begin{aligned}
\boldsymbol{E}\left\{N^s(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{y})\right\} &\doteq \begin{cases} \exp\{ns[\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)-R]\} & \hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)>R \\ \exp\{n[\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)-R]\} & \hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)\leq R \end{cases} && (74)\\
&= \exp\{n(s[\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)-R]-(1-s)[R-\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)]_+)\}, && (75)
\end{aligned}$$

Thus,

$$\begin{aligned}
& \boldsymbol{E}\left\{\left[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P(\boldsymbol{x}',\boldsymbol{y})\mathcal{I}[f(\boldsymbol{x}')=f(\boldsymbol{x})]\right]^s\right\} && (76)\\
\doteq \;& P^s(\boldsymbol{y})\sum_{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})} P^s(\boldsymbol{x}'|\boldsymbol{y})\exp\{n(s[\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)-R]-(1-s)[R-\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)]_+)\} && (77)\\
= \;& P^s(\boldsymbol{y})\sum_{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})} P^s(\boldsymbol{x}'|\boldsymbol{y})\exp\{n(s[\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)-R]-(1-s)[R-\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)]_+)\} && (78)\\
= \;& P^s(\boldsymbol{y})\sum_{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})} \exp\{-n(s[D(\hat{P}_{\boldsymbol{x}'|\boldsymbol{y}}\|P_{X|Y}|\hat{P}_{\boldsymbol{y}})+R]+(1-s)[R-\hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X|Y)]_+)\} && (79)\\
\doteq \;& P^s(\boldsymbol{y})\exp\left\{-n\min_{P_{X'|Y}}(s[D(P_{X'|Y}\|P_{X|Y}|\hat{P}_{\boldsymbol{y}})+R]+(1-s)[R-H(X'|Y)]_+)\right\} && (80)\\
\stackrel{\Delta}{=} \;& P^s(\boldsymbol{y})e^{-nL(\hat{P}_{\boldsymbol{y}},R,s)}, && (81)
\end{aligned}$$

where $\hat{P}_{\boldsymbol{x}'|\boldsymbol{y}}$ is the empirical conditional distribution of a random variable X' given Y induced by $(\boldsymbol{x}',\boldsymbol{y})$, and $D(P_{X'|Y}\|P_{X|Y}|P_Y)$ is defined as

$$D(P_{X'|Y}\|P_{X|Y}|P_Y)=\sum_y P_Y(y)\sum_x P_{X'|Y}(x|y)\log\frac{P_{X'|Y}(x|y)}{P_{X|Y}(x|y)}. \quad (82)$$

Consequently,

$$\begin{aligned}
\overline{\Pr\{\mathcal{E}_1\}} &\leq e^{nsT} \sum_{\boldsymbol{x},\boldsymbol{y}} P^{1-s}(\boldsymbol{x},\boldsymbol{y}) P^s(\boldsymbol{y}) e^{-nL(\hat{P}_{\boldsymbol{y}},R,s)} && (83)\\
&= e^{nsT} \sum_{\boldsymbol{y}} P(\boldsymbol{y}) e^{-nL(\hat{P}_{\boldsymbol{y}},R,s)} \sum_{\boldsymbol{x}} P^{1-s}(\boldsymbol{x}|\boldsymbol{y}) && (84)\\
&= e^{nsT} \sum_{\boldsymbol{y}} P(\boldsymbol{y}) e^{-nL(\hat{P}_{\boldsymbol{y}},R,s)} \prod_{i=1}^{n} \sum_{x\in\mathcal{X}} P^{1-s}(x|y_i) && (85)\\
&\doteq e^{-nE_1'(R,T,s)} && (86)
\end{aligned}$$

where

$$E_1'(R,T,s) = \min_{P_Y'} \left[D(P_Y'\|P_Y) + L(P_Y',R,s) - \sum_{y\in\mathcal{Y}} P_Y'(y) \ln \sum_{x\in\mathcal{X}} P^{1-s}(x|y) \right] - sT. \qquad (87)$$

Finally,

$$E_1'(R,T) = \sup_{s\geq 0} E_1'(R,T,s). \qquad (88)$$

Appendix A

Calculation of $L(R,s)$. Let $p_s = p^s/[p^s+(1-p)^s]$. Consider first the case $s\in[0,1]$, where $p_s \geq p$. In this case, the minimizer δ^* that achieves $L(R,s)$ is given by

$$\delta^* = \begin{cases} p & R < h(p) \\ h^{-1}(R) & h(p) \leq R < h(p_s) \\ p_s & R \geq h(p_s) \end{cases} \qquad (A.1)$$

Here, for $R < h(p)$, the derivative of the objective function vanishes only at $\delta = p > h^{-1}(R)$, where the term $[R-h(\delta)]_+$ vanishes. On the other hand, for $R \geq h(p_s)$, the derivative vanishes only at $\delta = p_s < h^{-1}(R)$, where the term $[R-h(\delta)]_+$ is active. In the intermediate range, the derivative jumps from a negative value to a positive value at $\delta = h^{-1}(R)$ discontinuously, hence it is a minimum. Thus, for $0\leq s\leq 1$, we have:

$$L(R,s) = \begin{cases} s[p\ln\frac{1-p}{p} + R - h(p)] & R < h(p) \\ sh^{-1}(R)\ln\frac{1-p}{p} & h(p) \leq R < h(p_s) \\ sp_s\ln\frac{1-p}{p} + R - h(p_s) & R \geq h(p_s) \end{cases} \qquad (A.2)$$

For $s > 1$, $p_s < p$. and so $h(p_s) < h(p)$. Here, for $R < h(p_s)$, which means also $R < h(p)$, the derivative vanishes only at $\delta = p > h^{-1}(R)$. On the other hand, for $R > h(p) > h(p_s)$, the

derivative vanishes only at $\delta = p_s < h^{-1}(R)$. In the intermediate range, $h(p_s) \le R < h(p)$, the derivative vanishes both at $\delta = p$ and $\delta = p_s$, so the minimum is the smaller between the two. Namely, it is $\delta^* = p_s$ if

$$sp_s \ln \frac{1-p}{p} + s[R - h(p_s)] + (1-s)[R - h(p_s)]_+ \le sp \ln \frac{1-p}{p} + s[R - h(p)] + (1-s)[R - h(p)]_+$$

or equivalently,

$$sp_s \ln \frac{1-p}{p} + R - h(p_s) \le sp \ln \frac{1-p}{p} + s[R - h(p)],$$

and it is $\delta = p^*$ otherwise. The choice between the two depends on R. Let

$$R(s) = \frac{s(p_s - p)\ln[(1-p)/p] + sh(p_s) - h(p)}{s-1} = -\frac{\ln[p^s + (1-p)^s]}{s-1} \tag{A.3}$$

Then, for $s > 1$,

$$L(R,s) = \begin{cases} s[p \ln \frac{1-p}{p} + R - h(p)] & R < R(s) \\ sp_s \ln \frac{1-p}{p} + R - h(p_s) & R \ge R(s) \end{cases} \tag{A.4}$$

Appendix B

Calculations of Error Exponents for Very Weakly Correlated BSS's. For $p = 1/2 - \epsilon$, we have, to the second order in ϵ, $H(X|Y) = h(p) = h(1/2-\epsilon) = \ln 2 - 2\epsilon^2$. Consider the range of rates $\ln 2 - 2\epsilon^2 < R \le \ln 2$. A second order Taylor series expansion of $\gamma(t) \stackrel{\Delta}{=} -\ln[(1/2-\epsilon)^t + (1/2+\epsilon)^t]$ around $\epsilon = 0$ (for fixed t) gives

$$\gamma(t) = (t-1)(\ln 2 - 2t\epsilon^2), \tag{B.1}$$

and so,

$$\begin{aligned} E_0(\rho, s) &= \gamma(1-s) + \rho\gamma\left(\frac{s}{\rho}\right) && \text{(B.2)} \\ &= -s[\ln 2 - 2(1-s)\epsilon^2] + (s-\rho)\left(\ln 2 - \frac{2s\epsilon^2}{\rho}\right) && \text{(B.3)} \\ &= 4s\epsilon^2 - 2s^2\left(1 + \frac{1}{\rho}\right)\epsilon^2 - \rho \ln 2. && \text{(B.4)} \end{aligned}$$

Now,

$$E_1(R,T) = \max_{0 \le s \le \rho \le 1} \left[s(4\epsilon^2 - T) - \rho(\ln 2 - R) - 2s^2\epsilon^2\left(1 + \frac{1}{\rho}\right)\right]. \tag{B.5}$$

We will find it convenient to present $R = \ln 2 - 2\theta^2\epsilon^2$, where $\theta \in [0,1]$, and so, from here on, the rate is parametrized by θ. The maximization over $\rho \geq s$, for a given s, is readily found to give

$$\rho_s^* = s|\epsilon|\sqrt{\frac{2}{\ln 2 - R}} = \frac{s}{\theta} \geq s, \tag{B.6}$$

On substituting $\rho = \rho_s^*$, we get

$$\begin{aligned}
E_1(R,T) &\leq \max_{0\leq s\leq 1}[E_0(\rho_s^*, s) + \rho_s^* R - sT] && \text{(B.7)}\\
&= \max_{0\leq s\leq 1}\left[s(4\epsilon^2 - T) - s|\epsilon|\sqrt{2(\ln 2 - R)} - 2s^2\epsilon^2 - 2s|\epsilon|\sqrt{\frac{\ln 2 - R}{2}}\right] && \text{(B.8)}\\
&= \max_{0\leq s\leq 1}\{s[4\epsilon^2 - T - 2|\epsilon|\sqrt{2(\ln 2 - R)}] - 2s^2\epsilon^2\} && \text{(B.9)}\\
&= \max_{0\leq s\leq 1}\{s[4\epsilon^2(1-\theta) - T] - 2s^2\epsilon^2\} && \text{(B.10)}
\end{aligned}$$

where the inequality is because when we maximized over ρ, we have ignored the constraint $\rho \leq 1$. Next, let $T = -\tau\epsilon^2$ for $\tau > 4$, then $s^* = 1$ and so,

$$E_1(R,T) \leq 4\epsilon^2(1-\theta) + \tau\epsilon^2 - 2\epsilon^2 = 2\epsilon^2(1-2\theta) + \tau\epsilon^2 \leq (\tau+2)\epsilon^2. \tag{B.11}$$

On the other hand,

$$\begin{aligned}
E_1'(R,T) &\geq \sup_{s\geq 1}[R - sT + \gamma(s) + \gamma(1-s)] && \text{(B.12)}\\
&= \sup_{s\geq 1}[s(4\epsilon^2 - T) - 4s^2\epsilon^2] + R - \ln 2 && \text{(B.13)}\\
&= \sup_{s\geq 1}[s(4\epsilon^2 - T) - 4s^2\epsilon^2] - 2\theta^2\epsilon^2 && \text{(B.14)}\\
&= \frac{(4\epsilon^2 - T)^2}{16\epsilon^2} - 2\theta^2\epsilon^2 && \text{(B.15)}\\
&\geq \frac{[(\tau+4)\epsilon^2]^2}{16\epsilon^2} - 2\epsilon^2 && \text{(B.16)}\\
&= \left[\frac{\tau(\tau+8)}{16} - 1\right]\epsilon^2. && \text{(B.17)}
\end{aligned}$$

References

[1] I. Csiszár, "Linear codes for sources and source networks: error exponents, universal coding," *IEEE Trans. Inform. Theory*, vol. IT–28, no. 4, pp. 585–592, July 1982.

[2] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inform. Theory*, vol. IT–26, no. 2, pp. 155–165, March 1980.

[3] I. Csiszár and J. Körner, "Graph decomposition: a new key to coding theorems," *IEEE Trans. Inform. Theory*, vol. IT–27, no. 1, pp. 5–12, January 1981.

[4] 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.

[5] S. C. Draper and E. Martinian, "Compound conditional source coding, Slepian–Wolf list decoding, and applications to media coding,"

[6] A. W. Eckford and W. Yu, "Rateless Slepian–Wolf codes," *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 1757–1761, 2005.

[7] 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.

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

[9] R. G. Gallager, "Source coding with side information and universal coding," LIDS-P-937, M.I.T., 1976.

[10] B. G. Kelly and A. B. Wagner, "Improved source coding exponents via Witsenhausen's rate," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 5615–5633, September 2011.

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

[12] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford University Press, New York 2009.

[13] Y. Oohama and T. S. Han, "Universal coding for the Slepian–Wolf data compression system and the strong converse theorem," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1908–1919, November 1994.

[14] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT–19, no. 4, pp. 471–480, January 1973.

[15] A. Somekh–Baruch and N. Merhav, "Exact random coding error exponents for erasure decoding," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6444–6454, October 2011.