Achievable Error Exponents in Multiterminal Source Coding

Krishnan Eswaran, Michael Gastpar
Dept. of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720
E-mail: {keswaran, gastpar}@eecs.berkeley.edu

***Abstract*— Encoding correlated sources at separate encoders has been studied extensively from the perspective of asymptotically long block codes. The associated error exponents are known for the case of lossless source coding. In this paper, we introduce a novel technique for deriving achievable error exponents for lossy source coding problems, where the original sources need to be reconstructed to within some fidelity. As an example, we show how to apply our technique to determine achievable error exponents for the Berger-Yeung problem.**

I. Introduction

Since the pioneering work of Slepian and Wolf [1], distributed or multiterminal source coding has been studied extensively. Thanks to new converse techniques, the sum-rate-distortion functions, and in some cases rate regions, of previously open multiterminal source coding problems have been completely characterized [2], [3]. These results have concerned lossy source coding problems, in which one is interested in compressing a source or sources with some fidelity.

In lossy distributed source coding, a standard way to show a rate region is achievable is to divide the procedure into two stages. In the first, we vector quantize the observations and in the second use a technique known as binning to further reduce the rate. The latter stage is due to an insight by Slepian and Wolf, who used a random hashing argument to show that a decoder can invert the hashing functions with high probability. The rate regions reflect the asymptotic performance of codes as the blocklength increases. In practice, however, large blocklengths introduce complexity and delay. Thus, the performance of such codes for finite blocklengths are also of interest.

Haroutunian introduced the problem of source coding exponents in lossy multiterminal source coding problems [8]. He then presented exponents for systems considered by Kaspi and Berger [9], as well as by Berger and Yeung [10]. Unfortunately, his results did not take into account the binning stage, so the exponents for the rest of the rate region remains open [11]. For the lossless source coding problem, error exponents have been studied to address these second order issues [4] [5] [6]. These studies amount to a study of the performance of the binning strategy if no quantization were done.

The difficulty in giving exponents (let alone tight ones) is how to interface between these two stages. When exponents are not a concern, there are two approaches to generate an interface. The first is a "super letter" coding strategy [12], in which blocks of the outputs from the quantization stage are treated as single source symbols and collected together for binning. The second approach, introduced by Berger [13] and Tung [14], is called the Markov lemma approach, in which one shows the output of the quantization stage is jointly typical with the source or sources, and then uses the properties of typical sequences to argue that the binning stage will be successful.

In this work, we analyze the second approach and present a novel technique for getting achievable exponents in lossy multiterminal source coding problems. We first review the Markov lemma, which gives an asymptotic result on the transitive property of typical sequences. Previous proofs have not been concerned with the convergence rate, and therefore have only shown that the convergence decays at least inversely with the blocklength [13]. We derive a bound on the rate of convergence of the Markov lemma. We show that it decays at least exponentially with the blocklength. As a result, our approach gives achievable exponents for all points achievable by the Berger-Tung inner bound [13], [14]. As an application, we present achievable exponents for the problem considered by Berger and Yeung [10].

II. Definitions and Notation

We begin by defining the notion of typicality we will use. All random variables are discrete and are represented by capital letters X, Y, Z, etc. Each random variable takes values in a finite set $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$. The cardinality of a set $\mathcal{X}$ is denoted $|\mathcal{X}|$. To simplify notation, we sometimes shorten the Cartesian product of two sets $\mathcal{X}$,$\mathcal{Y}$ from $\mathcal{X} \times \mathcal{Y}$ to $\mathcal{X}\mathcal{Y}$. Finally, for $x \in \mathcal{X}$ and a random vector X^n, $N(x; X^n)$ refers to the number of times the symbol x appears in the sequence. That is $N(x; X^n) = \sum_i \mathbf{1}_{\{X_i = x\}}$. Although it is a slight abuse of notation, for random variables X, Y, we use $p(x, y)$ to mean $P(X = x, Y = y)$.

Definition 1: Consider a distribution $p(\cdot, \cdot)$. A pair of sequences (x^n, y^n) are jointly δ-typical (denoted $(x^n, y^n) \in T_n(\delta)$) if for all $(a, b) \in \mathcal{X}\mathcal{Y}$ they satisfy

$$n^{-1}|N(a, b; x^n, y^n) - np(a, b)| < \frac{\delta}{|\mathcal{X}\mathcal{Y}|} \tag{1}$$

This is often called strong typicality in the literature [13].

Some of our error exponents will involve the binary Kullback-Leibler divergence.

Definition 2: The binary Kullback-Leibler divergence between two binary probability distributions $(p, 1-p)$ and $(q, 1-q)$ is

$$D(p\|q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}. \tag{2}$$

The following property about jointly δ-typical sequences will prove useful.

Lemma 1: Consider a function $g^n(x^n, y^n) = \frac{1}{n}\sum_{i=1}^{n} g(x_i, y_i)$. If X, Y have law $p(\cdot,\cdot)$ and $(x^n, y^n) \in T_n(\delta)$, then

$$g^n(x^n, y^n) \le Eg(X,Y) + \delta \max_{a,b} g(a,b). \tag{3}$$

Proof: The proof follows by reexpressing $g^n(x^n, y^n)$ in terms of our counting functions and applying the definition of δ-typical sequences. Doing so gives

$$\begin{aligned} g^n(x^n, y^n) &= \frac{1}{n}\sum_{i=1}^{n} g(x_i, y_i) && (4)\\ &= \sum_{a,b} g(a,b)\frac{N(a,b;x^n,y^n)}{n} && (5)\\ &\le \sum_{a,b} g(a,b)\left(p(a,b) + \frac{\delta}{|\mathcal{X}\mathcal{Y}|}\right) && (6)\\ &= Eg(X,Y) + \delta\max_{a,b} g(a,b). && (7) \end{aligned}$$

■

III. Markov Lemma

Typical sequences have been used to prove achievable results in both source and channel coding. There are two common notions of typical sequences. One notion of typicality has been to assume that the empirical entropy rate of a sequence is "close" to an actual entropy rate for some distribution on the sequences. For this version, a version of the weak law of large numbers show that random sequences that are stationary and ergodic become typical with high probability as the sequence length increases, which is called the asymptotic equipartition property. The second notion of typicality is to assume that the empirical probabilities match the probabilities for some model. A pair of sequences that satisfy such a property on their joint distribution are called jointly typical. One can show that pairs of sequences generated by i.i.d. distributions are jointly typical with high probability as the sequence length increases.

In distributed systems, one would like typical sequences generated at different sites to be jointly typical with each other. However, the following example, taken from [13], shows that typicality is not transitive.

Example 1: Let x^n, y^n, and z^n each be n-length binary sequences. Suppose we choose to model each as an i.i.d. sequence such that the joint probability at any index is $p(x,y,z) = \frac{1}{8}$. Then, we construct an example in which both (x^n, y^n) and (y^n, z^n) are each jointly typical, but (x^n, z^n) is not. Let x^n be a sequence of alternating 0s and 1s, y^n a sequence of alternating 00s and 11s, and $z^n = x^n$. Ignoring integer effects, $(1,1)$, $(1,0)$, $(0,1)$, and $(0,0)$ show up equal number of times in (x^n, y^n) and (y^n, z^n), so these are always jointly typical for $n = 4k$. Since $(0,1)$ and $(1,0)$ never show up in (x^n, z^n), it is not jointly typical for an appropriately chosen distance between the empirical and model probabilities.

One might notice that the problem with the above example is that the above sequences can be arbitrary and need not conform to the probability model given. Returning to that example, if x^n and y^n were realizations of random vectors X^n and Y^n observing the structure of the above model, standard arguments show that given that a sequence (Y^n, z^n) were jointly typical, (X^n, z^n) would be typical with high probability as the sequence length increases. In fact, this holds for any situation in which $p(x,y,z) = p(x)p(y|x)p(z|y)$. For this reason, the above fact is known as the Markov lemma. However, known arguments have only shown that probability (X^n, z^n) is not typical decays as n^{-1}.

A. Berger's Markov Lemma

Berger introduces the Markov lemma in [13] as a condition under which joint typicality is transitive. Here is a variation of Berger's version. Our statement is identical to his if we were to allow $\epsilon = 0$. For this case, however, Berger only shows that the probability of not being jointly typical decays inversely with the blocklength. Given the extra slack of $\epsilon > 0$, we show that the probability of not being jointly typical goes down exponentially in the blocklength.

Lemma 2: Suppose $X \leftrightarrow Y \leftrightarrow Z$ and that $\underline{X} = (X_1, \ldots, X_n)$ and $\underline{Y} = (Y_1, \ldots, Y_n)$ are such that each (X_k, Y_k) is distributed independently of all others according to $p(x,y) = \sum_z p(x,y,z)$. Then for all $\delta, \epsilon > 0$,

$$\begin{aligned} &P\left(\underline{X},\underline{z} \in T_n(|\mathcal{X}|(\delta+\epsilon)) \middle| (\underline{Y},\underline{z}) \in T_n(\delta)\right)\\ &\qquad \ge 1 - 2|\mathcal{X}\mathcal{Z}|\exp\left\{-nD\left(\frac{\epsilon|\mathcal{Z}|^{-1}+1}{2}\middle\|\frac{1}{2}\right)\right\}. \end{aligned} \tag{8}$$

Proof: Following identical arguments to Berger, we get that

$$\begin{gathered} np(x,z) - \delta|\mathcal{Z}|^{-1} \\ < E[N(x,z;\underline{X},\underline{z})|(\underline{Y},\underline{z}) \in T_n(\delta)] < \\ np(x,z) + \delta|\mathcal{Z}|^{-1}. \end{gathered}$$

Now, let $\underline{\tilde{X}}$ be such that $\tilde{X}_k = X_k$ for $k \neq i$ and $\tilde{X}_i$ is an independent copy of X_i. Then, for all i, x, z, we have that

$$|N(x,z;\underline{X},\underline{z}) - N(x,z;\underline{\tilde{X}},\underline{z})| \le 1,$$

almost surely. Since we have satisfied the assumptions necessary for [15, eq. (2.4.16)] to hold, we have

$$\begin{aligned} &P\left(N(x,z;\underline{X},\underline{z}) - np(x,z) \ge n(\delta+\epsilon)|\mathcal{Z}|^{-1} \middle| (\underline{Y},\underline{z}) \in T_n(\delta)\right)\\ &\le \exp\left\{-nD\left(\frac{\epsilon|\mathcal{Z}|^{-1}+1}{2}\middle\|\frac{1}{2}\right)\right\}. \end{aligned} \tag{9}$$

By symmetry we have that

$$P\left(|N(x,z;\underline{X},\underline{z})-np(x,z)|\geq n(\delta+\epsilon)|\mathcal{Z}|^{-1}\middle|(\underline{Y},\underline{z})\in T_n(\delta)\right)$$
$$\leq 2\exp\left\{-nD\left(\frac{\epsilon|\mathcal{Z}|^{-1}+1}{2}\middle\|\frac{1}{2}\right)\right\}. \tag{10}$$

Invoking the definition of joint typicality and applying the union bound gives us

$$P(\underline{X},\underline{z}\in T_n(|\mathcal{X}|(\delta+\epsilon))|(\underline{Y},\underline{z})\in T_n(\delta))$$
$$\leq 2|\mathcal{X}\mathcal{Z}|\exp\left\{-nD\left(\frac{\epsilon|\mathcal{Z}|^{-1}+1}{2}\middle\|\frac{1}{2}\right)\right\}. \tag{11}$$

The lemma follows immediately. ■

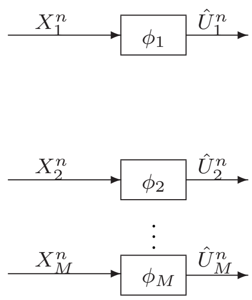

Fig. 1. Generalized Markov Lemma

B. Generalized Markov Lemma

While the above result is useful in the Wyner-Ziv problem, it has limited application elsewhere. We now present a generalized Markov lemma that turns out to be useful in proving achievability results for general multiterminal source coding problems. Berger [13] and Tung [14] call their version the extended Markov lemma. Housewright [12] shows a similar type of Markov lemma even though he does not describe it in those terms, and Oohama [16] shows a version for the Gaussian case. The following version most closely resembles the one proved by Han and Kobayashi [17].

Once again, our primary goal is to establish that the probability joint typicality is not satisfied goes down exponentially with the blocklength. The only difference is that now, we are extending our setup to multiple encoders, as seen in Figure 1. We now state the result.

Theorem 1: Let U_k be discrete auxiliary random variables with the property that

$$P(X_1=x_1,U_1=u_1,\ldots,X_M=x_M,U_M=u_M)=$$
$$P(X_1=x_1,\ldots,X_M=x_M)\prod_{k=1}^{M}P(U_k=u_k|X_k=x_k) \tag{12}$$

Let $\{\tilde{U}_k^n(l)\}_{l=1}^{L_k}$ be a set of length n random vectors, where the $\tilde{U}_k^n(l)$ are mutually independent random vectors of length n with each component i.i.d. according to the distribution $p(u_k)$. Then there exist functions $\phi_k:\mathcal{X}_k^n\rightarrow\{1,\ldots,L_k\}$ such that, for $\hat{U}_k^n=\tilde{U}_k^n\left(\phi_k(X_k^n)\right)$,

$$P((X_1^n,\ldots,X_M^n,\hat{U}_1^n,\ldots,\hat{U}_M^n)\notin T_n(\delta))$$
$$\leq 2|\mathcal{X}_1\cdots\mathcal{U}_M|\exp\left\{-nD\left(\frac{\delta/|\mathcal{X}_1\cdots\mathcal{U}_M|+1}{2}\middle\|\frac{1}{2}\right)\right\}$$
$$+\sum_{i=1}^{M}\exp\{-n(E_0(\rho_i,p(X_i,U_i))-\rho_iR_i)\}, \tag{13}$$

for $-1<\rho_i<0$. Further, when $R_k=\frac{\log L_k}{n}>I(X_k;U_k)$, there exists $\rho_k\in(-1,0)$ for which $E_0(\rho_k,p(X_k,U_k))-\rho_kR_k$ is positive.

Proof: The following observation about sequences of the auxiliary random variables is central to our proof.

Lemma 3: Let $(X_1(i),\ldots,X_M(i),U_1(i),\ldots,U_M(i))$ be i.i.d. in i with the joint distribution given in (12). Then

$$P((X_1^n,\ldots,X_M^n,U_1^n,\ldots,U_M^n)\notin T_n(\delta))$$
$$\leq 2|\mathcal{X}_1...\mathcal{U}_M|\exp\left\{-nD\left(\frac{\delta/|\mathcal{X}_1\cdots\mathcal{U}_M|+1}{2}\middle\|\frac{1}{2}\right)\right\}. \tag{14}$$

Proof: To simplify notation, define $Z(i)=(X_1(i),\ldots,X_M(i),U_1(i),\ldots,U_M(i))$. Note that they are i.i.d. Then $Z^n=(X_1^n,\ldots,X_M^n,U_1^n,\ldots,U_M^n)$. Now, let $\tilde{Z}^n$ be such that $\tilde{Z}(k)=Z(k)$ for $k\neq i$, and $\tilde{Z}(i)$ is an independent copy of $Z(i)$. Then, we know that for all $\mathbf{a}\in|\mathcal{X}_1\cdots\mathcal{U}_M|$ and all i,

$$|N(\mathbf{a};Z^n)-N(\mathbf{a};\tilde{Z}^n)|\leq 1$$

almost surely. Then the Azuma-Hoeffding-Bennett inequalities [15, eq. (2.4.16)] imply that for all $\mathbf{a}\in\mathcal{X}_1\cdots\mathcal{U}_M$ and $\beta=\delta/|\mathcal{X}_1\cdots\mathcal{U}_M|$

$$P(|N(\mathbf{a};X_1^n,\ldots,X_M^n,U_1^n,\ldots,U_M^n)-np(\mathbf{a})|\geq n\beta)$$
$$\leq 2\exp\left\{-nD\left(\frac{\beta+1}{2}\middle\|\frac{1}{2}\right)\right\}. \tag{15}$$

By the definition of joint δ-typicality and applying the union bound, we get (14). ■

The remaining terms come from the fact that our functions need to approximate this structure. We can show how this works using a forbidden codeword argument, as seen in [18] and [12].

The forbidden codeword idea is as follows. Let each source X_k pass through a discrete memoryless channel (DMC)

$$P(U_k=u_k|X_k=x_k), \tag{16}$$

based on the joint probability given in (12). Suppose that, in addition to source X_k, Encoder$-k$ also has access to the output of the DMC X_k passes through. If this channel output were an additional codeword in our codebook, we would always select it, and joint typicality would follow immediately from Lemma 3. However, this codeword is forbidden because the codebook must be defined before the source can be observed. Thus, we will show that we can select codewords from our original codebooks that are at least as likely to be jointly δ-typical as the forbidden codewords. We will find the probability that the forbidden codewords are more likely to be jointly δ-typical is less than

$$\sum_{i=1}^{M} \exp\{-n(E_0(\rho_i, p(X_i, U_i)) - \rho_i R_i)\}. \tag{17}$$

Thus, the probability that the functions ϕ_k will not yield sequences that are jointly δ-typical with the X_k^n sequences will be upper bounded by the sum of (14) and (17).

We now specify functions ϕ_k that result in this. To simplify notation, we define

$$\begin{aligned}\gamma_k(\tilde{u}_k^n(l)) = P((&X_1^n, \ldots, X_M^n, U_1^n, \ldots, U_{k-1}^n, \\ &\tilde{u}_k^n(l), \hat{u}_{k+1}^n, \ldots, \hat{u}_M^n) \notin T_n(\delta) | X_k^n),\end{aligned} \tag{18}$$

which is an X_k^n-measurable random variable. This is a slight abuse of notation since for $\gamma_k(U_k^n)$, the expectation would be taken over U_k^n. Using this notation, our functions ϕ_k are defined as

$$\phi_k(X_k^n) = \arg \min_{l: 1 \leq l \leq L_k} \gamma_k(\tilde{u}_k^n(l)) \tag{19}$$

With this definition for our encoder functions, we now state the following lemma.

Lemma 4: For the ϕ_k defined as in (19) and $\hat{u}_k^n = \phi_k(X_k^n)$,

$$\begin{aligned} & P((X_1^n, \ldots, X_M^n, \hat{u}_1^n, \ldots, \hat{u}_M^n) \notin T_n(\delta)) \\ \leq \; & P((X_1^n, \ldots, X_M^n, U_1^n, \ldots, U_M^n) \notin T_n(\delta)) \\ & + \sum_{i=1}^{M} P(\gamma_k(\hat{u}_k^n) > \gamma_k(U_k^n)) \end{aligned} \tag{20}$$

Proof: Observe that

$$\begin{aligned} & P((X_1^n, \ldots, X_M^n, \hat{u}_1^n, \ldots, \hat{u}_M^n) \notin T_n(\delta)) \\ & \quad = E[\gamma_1(\hat{u}_1^n)] && (21) \\ = \; & E[\gamma_1(\hat{u}_1^n) \mathbf{1}_{\{\gamma_1(\hat{u}_1^n) \leq \gamma_1(U_1^n)\}}] \\ & \quad + E[\gamma_1(\hat{u}_1^n) \mathbf{1}_{\{\gamma_1(\hat{u}_1^n) > \gamma_1(U_1^n)\}}] && (22) \\ \leq \; & E[\gamma_1(U_1^n) \mathbf{1}_{\{\gamma_1(\hat{u}_1^n) \leq \gamma_1(U_1^n)\}}] \\ & \quad + E[\gamma_1(\hat{u}_1^n) \mathbf{1}_{\{\gamma_1(\hat{u}_1^n) > \gamma_1(U_1^n)\}}] && (23) \\ \leq \; & E[\gamma_1(U_1^n)] + E[\mathbf{1}_{\{\gamma_1(\hat{u}_1^n) > \gamma_1(U_1^n)\}}] && (24) \\ = \; & E[\gamma_2(\hat{u}_2^n)] + P(\gamma_1(\hat{u}_1^n) > \gamma_1(U_1^n)), && (25) \end{aligned}$$

where (21) follows from successive conditioning, (22) by linearity of expectation, (23) from the event on the indicator function, (24) since the indicator function and γ_1 are bounded by 1 almost surely, and (25) from the definition of $\gamma_2(\cdot)$. The result follows by induction since we can apply the same arguments to show that

$$E[\gamma_k(\hat{u}_k^n)] \leq E[\gamma_k(\hat{u}_{k+1}^n)] + P(\gamma_k(\hat{u}_k^n) > \gamma_k(U_k^n)) \tag{26}$$

and noting that $E[\gamma_n(U_n^n)]$ is the same as the left hand side of (14). ■

We now take the expectation over the codebook ensemble, denoting it E_C. Since we have the bound (14), all we need is to show that

$$P(\gamma_k(\hat{U}_k^n) > \gamma_k(U_k^n)) \leq \exp\{-n(E_0(\rho_i, p(X_i, U_i)) - \rho_i R_i)\} \tag{27}$$

and that these go to 0 when $R_k = \frac{\log L_k}{n} > I(X_k; U_k)$. The latter is well known (see e.g. [19], [18], [12]), and the former is shown below.

$$\begin{aligned} & E_C P(\gamma_k(\hat{u}_k^n) > \gamma_k(U_k^n)) \\ & \quad = \sum_{\substack{\tilde{u}_k^n(l'), \\ 1 \leq l' \leq L_k}} \prod_{l=1}^{L_k} p^n(\tilde{u}_k^n(l)) \sum_{x^n, u_k^n} p^n(u_k^n, x_k^n) \mathbf{1}_{A_k} && (28) \\ & \quad = \sum_{\substack{\tilde{u}_k^n(l'), \\ 1 \leq l' \leq L_k}} \prod_{l=1}^{L_k} p^n(\tilde{u}_k^n(l)) \sum_{x^n, u_k^n} p^n(u_k^n) p^n(x_k^n | u_k^n) \mathbf{1}_{A_k}, && (29) \end{aligned}$$

where $A_k = \{\gamma_k(u_k^n) \leq \gamma_k(\tilde{u}_k^n(m)), \forall m\}$. For fixed x_k^n, Hölder's inequality implies that

$$\begin{aligned} & E_C P(\gamma_k(\hat{u}_k^n) > \gamma_k(U_k^n)) \\ & \leq \sum_{x^n} \sum_{\substack{\tilde{u}_k^n(l'), \\ 1 \leq l' \leq L_k}} \prod_{l=1}^{L_k} p^n(\tilde{u}_k^n(l)) \\ & \qquad \cdot \frac{\left(\sum_{u_k^n} p^n(u_k^n) p^n(x_k^n | u_k^n)^{1/(1+\rho)} \right)^{1+\rho}}{\left(\sum_{u_k^n} p^n(u_k^n) \mathbf{1}_{A_k} \right)^{\rho}}, \end{aligned} \tag{30}$$

for $-1 < \rho < 0$. Since $(\cdot)^{-\rho}$ in convex-$\cap$, we have

$$\begin{aligned} & E_C P(\gamma_k(\hat{u}_k^n) > \gamma_k(U_k^n)) \\ & \leq \frac{\sum_{x^n} \left(\sum_{u_k^n} p^n(u_k^n) p^n(x_k^n | u_k^n)^{1/(1+\rho)} \right)^{1+\rho}}{\left(\sum_{\substack{\tilde{u}_k^n(l'), \\ 1 \leq l' \leq L_k}} \prod_{l=1}^{L_k} p^n(\tilde{u}_k^n(l)) \sum_{u_k^n} p^n(u_k^n) \mathbf{1}_{A_k} \right)^{\rho}}. \end{aligned} \tag{31}$$

Since the marginals are all the same, symmetry allows us to

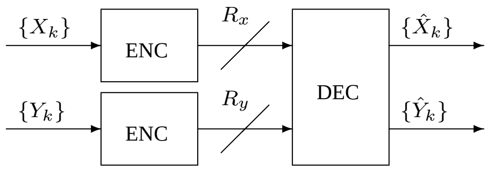

Fig. 2. Berger-Yeung Problem

simplify this expression to

$$E_C P(\gamma_k(\hat{u}_k^n) > \gamma_k(U_k^n))$$

$$\leq \sum_{x^n} \left(\sum_{u_k^n} p^n(u_k^n) p^n(x_k^n|u_k^n)^{1/(1+\rho)} \right)^{1+\rho} \left(\frac{1}{L_k+1} \right)^{-\rho} \tag{32}$$

$$\leq \sum_{x^n} \left(\sum_{u_k^n} p^n(u_k^n) p^n(x_k^n|u_k^n)^{1/(1+\rho)} \right)^{1+\rho} \left(\frac{1}{L_k} \right)^{-\rho} \tag{33}$$

$$= \sum_{x^n} \left(\sum_{u_k^n} p^n(u_k^n) p^n(x_k^n|u_k^n)^{1/(1+\rho)} \right)^{1+\rho} e^{n\rho R_k} \tag{34}$$

$$= \left(\sum_{x} \left(\sum_{u_k} p(u_k) p(x_k|u_k)^{1/(1+\rho)} \right)^{1+\rho} \right)^{n} e^{n\rho R_k} \tag{35}$$

$$= \exp\{-n(E_0(\rho, p(X_k, U_k)) - \rho R_k)\}, \tag{36}$$

where $E_0(\rho, p(X_k, U_k)) = -\log\left(\sum_x \left(\sum_{u_k} p(u_k)p(x_k|u_k)^{1/(1+\rho)}\right)^{1+\rho}\right)$.

Finally, combining (14), (20), and (36) gives us (13). ■

IV. Application to Berger-Yeung Problem

We now consider a problem considered by Berger and Yeung [10], which has the structure depicted in Figure 2. In this problem, there are two correlated sources, $\{X_k\}$ and $\{Y_k\}$, at separate and non-interacting encoders. As the blocklength gets large, the decoder is interested in reconstructing $\{X_k\}$ perfectly and to approximate $\{Y_k\}$ to within some distortion D for a given distortion function d. For a distortion D, Berger and Yeung characterized the rate-region for this problem (R_x, R_y) as

1) $X \leftrightarrow Y \leftrightarrow Z$ for an auxiliary random variable $Z \in \mathcal{Z}$,
2) $\hat{Y}(X, Z)$ exists such that $Ed(Y, \hat{Y}) \leq D$,
3) $R_x \geq H(X|Z)$, $R_y \geq I(Y; Z|X)$, and $R_x + R_y \geq H(X) + I(Y; Z|X)$,
4) $|\mathcal{Z}| \leq |\mathcal{Y}| + 2$.

We are interested in the probability that a code fails to meet these requirements for any finite blocklength. That is, we are interested in the event $B(\alpha)$, where

$$B(\alpha)^c = \{d^n(Y^n, \hat{Y}^n) \leq D + \alpha, X^n = \hat{X}^n\}. \tag{37}$$

For interesting results, $\alpha > 0$.

Theorem 2: Let (R_x, R_y, D) be an achievable point in the region of the Berger-Yeung problem. Then, given any $\alpha, \epsilon > 0$, then there exists a sequence of Berger-Yeung codes with rate $(R_x, R_y) + (\epsilon, \epsilon)$ such that

$$\begin{aligned} &\liminf_{n\to\infty} -\frac{1}{n} \log P\left(B(\alpha)\right) \\ &\geq \max_{\substack{\delta \leq \min\{\alpha/D_{\max},\\ \epsilon/\max\{c_x, c_z\}\}}} \max_{p(z|y): Ed(Y,\hat{Y}) \leq D} \min \Bigg\{ \\ &R_x + R_y - H(X) - I(Y; Z|X) - c\delta, \\ &D\left(\frac{\delta/|\mathcal{XYZ}| + 1}{2} \middle\| \frac{1}{2} \right) \Bigg\}, \end{aligned} \tag{38}$$

where $c_x = \max_x \log \frac{1}{p(x)}$, $c_z = \max_z \log \frac{1}{p(z)}$, $c_{xz} = \max_{x,z} \log \frac{1}{p(x,z)}$, and $c = c_x + c_z + c_{xz}$.

Proof: Our encoding procedure consists of two phases: a quantization phase followed by a binning phase.In the quantization phase, the encoder for X^n does nothing to the source since we are interested in perfect reconstruction. The encoder for Y^n, on the other hand, follows the following quantization procedure. Using $e^{n\tilde{R}}$ codewords generated i.i.d. by distribution $p(z^n) = \prod p(z_i)$, we generate the output $Z^n(Y^n)$. By the generalized Markov lemma given in Theorem 1, we know that there exists an encoding rule such that the probability this output will not be jointly δ-typical with our sources has probability at most

$$\begin{aligned} &P((X^n, Y^n, Z^n(Y^n)) \notin T_n(\delta)) \\ &\leq 2|\mathcal{XYZ}| e^{-nD\left(\frac{\delta/|\mathcal{XYZ}|+1}{2} \middle\| \frac{1}{2}\right)} \\ &\quad + e^{-n(E_0(\rho, p(Y,Z)) - \rho\tilde{R})}, \end{aligned} \tag{39}$$

where $-1 < \rho < 0$. Note that we can make $\tilde{R}$ large enough so that the first term in (39) dominates. Further, if we let $\delta \leq \alpha/D_{\max}$, where $D_{\max} = \max_{y,\hat{y}} d(y, \hat{y})$, then this is the probability our target distortion is not met.

The second encoding phase is binning. In this stage, the encoder for X^n selects $\frac{e^{n(H(X)+\epsilon)}}{e^{nR_x}}$ δ-typical sequences of $p(x^n)$ independently and randomly for each of e^{nR_x} bins. It then finds the bin containing X^n or declares an error. For the encoder for Y^n, we select $\frac{e^{n(I(Y;Z)+\epsilon)}}{e^{nR_y}}$ typical sequences of $p(z^n)$ independently and randomly for each of e^{nR_y} bins. It then finds the bin containing $Z^n(Y^n)$ or declares an error. We require that $R_x \leq H(X) + \epsilon$ and $R_y \leq I(Y; Z) + \epsilon$ for the bins to be nonempty. Assuming there were no error in the quantization stage, the probability that each encoder declares an error is [13]

$$P(X^n \text{ not found}|X^n \in T_n(\delta)) \leq \exp\left\{-e^{n(\epsilon - c_x\delta)}\right\}, \tag{40}$$

$$\begin{aligned} &P(Z^n(Y^n) \text{ not found}|Z^n(Y^n) \in T_n(\delta)) \\ &\qquad \leq \exp\left\{-e^{n(\epsilon - c_z\delta)}\right\}, \end{aligned} \tag{41}$$

where $c_x = \max_x \log \frac{1}{p(x)}$, $c_z = \max_z \log \frac{1}{p(z)}$. Thus, we require $\epsilon > \delta \max\{c_x, c_z\}$.

The decoder receives the bin indices from the two encoders and must decipher which pair of codewords was sent by the encoders. If there is only one pair of codewords that are jointly typical in the product bin, then the decoder can identify them. Otherwise, it declares an error. Because each typical sequence was selected independently and randomly for the bins, a union bound argument gives a bound for the error probability as

$$P(E_2) \leq \frac{e^{n(H(X)+\epsilon)}}{e^{nR_x}} \cdot \frac{e^{n(I(Y;Z)+\epsilon)}}{e^{nR_y}} \cdot \frac{|x^n, z^n \in T_n(\delta)|}{|x^n \in T_n(\delta)||z^n \in T_n(\delta)|} \quad (42)$$

$$\leq e^{-n(R_x+R_y-H(X)-I(Y;Z|X)-2\epsilon-c\delta)} \quad (43)$$

since $|x^n \in T_n(\delta)| \geq e^{n(H(X)+\delta c_x)}$, $|z^n \in T_n(\delta)| \geq e^{n(H(Z)+\delta c_z)}$, $|x^n, z^n \in T_n(\delta)| \leq e^{n(H(X,Z)+\delta c_{xz})}$ and where $c = c_x + c_z + c_{xz}$, $I(Y;Z) = I(X,Y;Z)$, and where $c_{xz} = \max_{x,z} \log \frac{1}{p(x,z)}$.

The total error probability is bounded by the sum of (39), (40), (41), and (43). Making $\tilde{R}$ large enough so that the first term on the right-hand side of (39) dominates and noting that (40) and (41) are doubly exponential gives us our result. ■

V. Discussion

We have demonstrated a novel technique for deriving achievable source coding error exponents in multiterminal source coding problems. Whenever the Berger-Tung inner bound is tight, this approach gives a lower bound on the achievable exponent. However, the typical sequence approach appears to be a little coarse when compared to the more elegant results of lossless source coding. Berger gave some general results on this for the point-to-point problem [7, pp. 194-198]. Unfortunately, the binning argument precludes using such an argument in our case. Still, one would like to determine the tightest possible exponents for this problem. Another issue for future consideration is lower bounds to our error probability.

VI. Acknowledgments

The work presented in this paper was supported in part by the National Science Foundation under awards CCF-0347298 (CAREER) and CNS-0326503.

References

[1] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, pp. 471 – 480, 1973.

[2] A. B. Wagner, S. Tavildar, and P. Viswanath, “The rate region of the quadratic Gaussian two-terminal source-coding problem,” 2005. [Online]. Available: http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:cs/0510095

[3] S. Tavildar and P. Viswanath, “The sum rate for the vector Gaussian CEO problem,” in *2005 Asilomar Conference*, 2005.

[4] R. Gallager, “Source coding with side information and universal coding,” 1979, lIDS Technical Report.

[5] I. Csiszár and J. Körner, “Towards a general theory of source networks,” *IEEE Transactions on Information Theory*, vol. 26, p. 155, 1980.

[6] I. Csiszár and J. G. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic Press, Inc., 1982.

[7] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, ser. Information and System Sciences Series. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[8] E. A. Haroutunian, “Multiterminal source coding achievable rates and reliability,” *IEEE Transactions on Information Theory*, vol. 42, no. 6, p. 2094.

[9] A. H. Kaspi and T. Berger, “Rate-distortion for correlated sources with partially separated encoders,” *IEEE Transactions on Information Theory*, vol. IT-28, no. 6, p. 828, 1982.

[10] T. Berger and R. Yeung, “Multiterminal source encoding with one distortion criterion,” *IEEE Transactions on Information Theory*, vol. IT-35, p. 228, 1989.

[11] E. A. Haroutunian, “Acknowledgment,” *IEEE Transactions on Information Theory*, vol. 44, no. 7, p. 3177, 1998.

[12] K. B. Housewright, “Source coding studies for multiterminal systems,” Ph.D. dissertation, University of California, Los Angeles, 1977.

[13] T. Berger, “Multiterminal source coding,” in *Lecture Notes presented at CISM Summer School on the Information Theory Approach to Communications*, 1977.

[14] S. Tung, “Multiterminal source coding,” Ph.D. dissertation, Cornell University, 1977.

[15] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications, 2nd ed*, ser. Stochastic Modelling and Applied Probability. Springer, 1998.

[16] Y. Oohama, “The rate-distortion function for the quadratic Gaussian CEO problem,” *IEEE Transactions on Information Theory*, vol. 44, pp. 1057–1070, May 1998.

[17] T. S. Han and K. Kobayashi, “A unified achievable rate region for a general class of multiterminal source coding systems,” *IEEE Transactions on Information Theory*, pp. 277–288, May 26.

[18] J. Omura, “A coding theorem for discrete-time sources,” *IEEE Transactions on Information Theory*, vol. 19, pp. 490–498, 1973.

[19] R. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, 1968.