ISIT'2013 1569720579

A Technique for Deriving One-Shot Achievability Results in Network Information Theory

Mohammad Hossein Yassaee, Mohammad Reza Aref, Amin Gohari
Information Systems and Security Lab (ISSL),
Sharif University of Technology, Tehran, Iran,
E-mail: yassaee@ee.sharif.edu, {aref,aminzadeh}@sharif.edu.

***Abstract*—This paper proposes a novel technique to prove a one-shot version of achievability results in network information theory. The technique is not based on covering and packing lemmas. In this technique, we use a stochastic encoder and decoder with a particular structure for coding that resembles both the ML and the joint-typicality coders. Although stochastic encoders and decoders do not usually enhance the capacity region, their use simplifies the analysis. The Jensen inequality lies at the heart of error analysis, which enables us to deal with the expectation of many terms coming from stochastic encoders and decoders at once. The technique is illustrated via four examples: point-to-point channel coding, Gelfand-Pinsker, broadcast channel and Berger-Tung problem of distributed lossy compression. Applying the one-shot result for the memoryless broadcast channel in the asymptotic case, we get the entire region of Marton's inner bound without any need for time-sharing. Also, these results are employed in conjunction with multi-dimensional berry-esseen CLT to derive new regions for finite-blocklength regime of Gelfand-Pinsker.**

I. Introduction

Information theory aims to find optimal reliable communication rates in networks. The combinatorial structure of networks makes the problem difficult in general. However one can employ the law of large numbers by looking at asymptotic behavior of networks for large blocklengths. But this comes at the cost of a long delay. This motivates looking at the problem in the so called "finite blocklength regime." The blocklength in this regime is not infinitely long, but is sufficiently large for certain CLTs to hold. Originally studied by Strassen [1], there has been a recent surge of works on this topic following the results of Polyanskiy et al [2] (see for instance [3]-[5]).

In this paper we consider *one-shot* network information theory where a *single* use of the network is allowed. In this case the probability of error cannot always be driven to zero. Further, well-known techniques such as joint typicality and time sharing are not applicable here. Given an admissible probability of error, our goal is to find a characterization of a set of achievable rates that resembles the form of the asymptotic results. There has been some previous work along this direction. Wang and Renner [6] derive one-shot upper and lower bounds for the problem of transmission of classical information over a classic-quantum channel (see also [7]). Recently Verdu has proposed a one-shot version of the covering and packing lemmas, and has applied it to a set of classical problems in information theory [8].

This work was supported by Iran-NSF under grant No. 88114.46.

Our main contribution is a proof technique for deriving the results on the one-shot region. The technique uses elementary tools and is not based on extensions of packing or covering lemmas. It is based on a particular construction for encoder and decoders that is not ML, but resembles both the ML and jointly typical coders. Our proposed decoders are stochastic and intuitively attempt to pass the received symbol through a certain inverse conditional distribution. The Jensen's inequality is central to the analysis of the error. The technique can be widely applied to problems of network information theory. To illustrate this, we derive new results for the problems of Gelfand-Pinsker, broadcast channel and Berger-Tung problem of distributed lossy compression. The asymptotic forms of these expressions is the same as that of classical results. Our one-shot bounds also imply new results in the finite blocklength regime.

The most related previous work is that of Verdu [8]. Whereas [8] proposes a one-shot covering and packing lemmas to solve network problems, we propose a direct analysis comprising of a chain of inequalities. By bypassing the need for covering and packing lemmas, we can provide bounds for problems that were originally solved using mutual covering and packing lemmas in the asymptotic regime. This is helpful because no one-shot extension of the mutual covering and packing lemma exists. We discuss this point in more details in Remark 3.

This paper is organized as follows: In Section II we provide some definitions. This is followed by four sections that provide a one-shot result for the point-to-point channel, Gelfand-Pinsker, broadcast channel (Marton) and Berger-Tung problem. We also provide a finite blocklength result for the Gelfand-Pinsker. The finite blocklength results for the other cases are similar and can be found in [11].

II. Definitions

Definition 1: For a pmf $p_{X,Y,Z}$, the *conditional information density* $\imath(x;y|z)$ is defined by

$$\imath_p(x;y|z) := \log \frac{p(x,y|z)}{p(x|z)p(y|z)},$$

and for general r.v.'s it is defined by

$$\imath_p(x;y|z) := \log \frac{\mathsf{d}p_{X,Y|Z}}{\mathsf{d}(p_{X|Z} \times p_{Y|Z})}(x,y,z).$$

Whenever the underlying distribution is clear from the context, we drop the subscript p from $\imath_p(x;y|z)$.

Definition 2: Let $\mathbf{X}$ be a multi-dimensional normal variable with zero mean and covariance matrix V. The complementary multivariate Gaussian cumulative distribution region associated with V is defined by

$$\mathcal{Q}^{-1}(\mathsf{V},\epsilon) := \{\mathbf{x} : \mathsf{P}(\mathbf{X} \le \mathbf{x}) \ge 1-\epsilon\}.$$

We use M and J to denote size of alphabets of random variables M and J, respectively, i.e. $\mathsf{M} = |\mathcal{M}|$ and $\mathsf{J} = |\mathcal{J}|$.

III. Point-to-point channel

We begin our illustration of the one-shot achievability proof with the classical point-to-point channel. Consider a channel with the law $q_{Y|X}$ and an input distribution q_X. Let $\mathcal{C} = \{X(1), \cdots, X(\mathsf{M})\}$ be a random codebook where the elements $X(i)$ are drawn independently from q_X (each codeword $X(i)$ is only a single rv). As usual, $X(m)$ is the codeword used for transmission of the message m. For the decoding we use an stochastic variation of MAP decoding. Instead of declaring the message $\hat{m}$ with maximal posterior probability as in MAP, the decoder randomly draws a message $\hat{m}$ from the conditional pmf $P_{M|Y}$, where P is the induced pmf by the code, $P_{M,Y}(m,y) = \frac{1}{\mathsf{M}} q(y|X(m))$.[1] More specifically,

$$P_{M|Y}(\hat{m}|y) = \frac{q(y|X(\hat{m}))}{\sum_{\bar{m}} q(y|X(\bar{m}))} = \frac{2^{\imath_q(y;X(\hat{m}))}}{\sum_{\bar{m}} 2^{\imath_q(y;X(\bar{m}))}}. \quad (1)$$

The mutual information term $\imath_q(y;X(\hat{m}))$ is computed using pmf $q_X q_{Y|X}$ that has nothing to do with the pmf induced by the code. However the sequence $X(\hat{m})$ itself is random, hence we have used $P_{M|Y}(\hat{m}|y)$ (capital P) to denote the pmf.

We refer this decoder as *stochastic likelihood coder* (SLC), or as *stochastic mutual information coder* (SMC).[2] The second equality shows that the probability of selecting a message is proportional to two to the power of its mutual information with the received output. So codewords with higher mutual information have a higher chance of being selected as the output of the decoder. This resembles the widely used joint typicality decoder in the asymptotic regime.

Theorem 1: The *expected value* of the probability of *correct decoding* of SLC (or SMC) for a randomly generated codebook of size M is bounded from below by

$$\mathbb{E}_{\mathcal{C}}\mathsf{P}[C] \ge \mathbb{E}_{q_{XY}} \frac{1}{1+(\mathsf{M}-1)2^{-\imath_q(X;Y)}}. \quad (2)$$

Proof: Observe that the joint distribution of random variables factors as,

$$P_{MY\hat{M}}(m,y,\hat{m}) = \frac{1}{\mathsf{M}} q(y|X(m)) P_{M|Y}(\hat{m}|y),$$

and the probability of correct decoding can be written as $\mathsf{P}[C] = \sum_{m,y} P_{MY\hat{M}}(m,y,m)$, hence we have:

$$\mathbb{E}\mathsf{P}[C] = \mathbb{E}\sum_{m,y} \frac{1}{\mathsf{M}} q(y|X(m)) \frac{2^{\imath_q(y;X(m))}}{\sum_{\bar{m}} 2^{\imath_q(y;X(\bar{m}))}} \quad (3)$$

$$= \mathbb{E}\sum_{y} q(y|X(1)) \frac{2^{\imath_q(y;X(1))}}{\sum_{\bar{m}} 2^{\imath_q(y;X(\bar{m}))}} \quad (4)$$

$$= \sum_{y} \mathbb{E}_{X(1)}\mathbb{E}_{\mathcal{C}|X(1)} q(y|X(1)) \frac{2^{\imath_q(y;X(1))}}{\sum_{\bar{m}} 2^{\imath_q(y;X(\bar{m}))}} \quad (5)$$

$$\ge \sum_{y} \mathbb{E}_{X(1)} q(y|X(1)) \frac{2^{\imath_q(y;X(1))}}{\mathbb{E}_{\mathcal{C}|X(1)} \sum_{\bar{m}} 2^{\imath_q(y;X(\bar{m}))}} \quad (6)$$

$$= \sum_{y} \mathbb{E}_{X(1)} q(y|X(1)) \frac{2^{\imath_q(y;X(1))}}{2^{\imath_q(y;X(1))} + (\mathsf{M}-1)} \quad (7)$$

$$= \sum_{x,y} q(x)q(y|x) \frac{2^{\imath_q(y;x)}}{2^{\imath_q(y;x)} + (\mathsf{M}-1)} \quad (8)$$

$$= \mathbb{E}_{XY} \frac{1}{1+(\mathsf{M}-1)2^{-\imath_q(X;Y)}}, \quad (9)$$

where (5) follows from the law of iterated expectations, (6) follows from the Jensen inequality for the convex function $f(x) = x^{-1}$ on $\mathbb{R}^+$, and (7) follows from the following equation for $\bar{m} \neq 1$,

$$\mathbb{E}_{\mathcal{C}|X(1)} 2^{\imath(y;X(\bar{m}))} = \sum_x q(x) 2^{\imath(y;x)} = \sum_x q(x|y) = 1,$$

where we use the fact that $X(\bar{m})$ is independent of $X(1)$ for $\bar{m} \neq 1$ and is drawn from q_X. ∎

IV. Gelfand-Pinsker

Consider the problem of transmitting a message over a state-dependent channel with state information available at the encoder. Let q_S and $q_{Y|X,S}$ be the state's pmf and the channel transition probability, respectively.

A. One-shot achievability

Theorem 2: Given any $q_{U|S}$ and function $x(u,s)$, there is a code for a single use of the channel whose probability of correct decoding is bounded from below by

$$\mathbb{E}_{USY} \frac{1}{(1+\mathsf{J}^{-1}2^{\imath(U;S)})(1+\mathsf{MJ}2^{-\imath(U;Y)})}, \quad (10)$$

where $\mathsf{J} > 0$ is an arbitrary integer. Moreover, loosening this bound gives the following upper bound on the error probability of the code,

$$\mathsf{P}[\log \mathsf{J} - \imath(U;S) < \gamma \text{ , or } \imath(U;Y) - \log \mathsf{MJ} < \gamma] + 3 \times 2^{-\gamma}, \quad (11)$$

where γ is any positive number.

Remark 1: If we apply the above result to n copies of a memoryless state dependent channel, we recover the asymptotic Gelfand-Pinsker result. In this derivation the first term in the denominator of (10), $1+\mathsf{J}^{-1}2^{\imath(U;S)}$ corresponds to a covering lemma in the asymptotic case, while the second term $1+\mathsf{MJ}2^{-\imath(U;Y)}$ corresponds to a packing lemma. Observe that the first term is proportional to J^{-1} whereas the second term is proportional to MJ. Thus the above formula combines covering and packing lemmas at once.

Remark 2: If we further loosen the first term of eq. (11) using the union bound, we get Verdu's bound on this problem

[1] The pmf is random due to the random codebook.

[2] The reason for introducing two names for apparently the same object will become clear later. These decoders will not be the same in other problems.

[8] except for the term $3 \times 2^{-\gamma}$. This residual term is not of significance in the finite blocklength n-letter regime where we choose γ of the order of $\log(n)$ (see Theorem 3); the main contribution comes from the probability terms. In a concurrent work [9], Watanabe et.al., prove an expression similar to eq. (11) using a different method based on channel simulation. They also applied their approach to the problem of source coding with a helper and to the Wyner-Ziv problem. It is not clear whether their approach is applicable to the scenarios such as broadcast channel, multiple description coding, etc that is solved in the asymptotic case using the multivariate covering lemma, since no extension of channel simulation is known for multiuser scenarios. Nonetheless, our technique bypasses the need for either an extension of covering lemma to multivariate covering, or a multi-terminal extension of the channel simulation result.

Proof: We only prove the lower bound on probability of correct decoding. Derivation of the loosened bound can be found in [11]. Let $\mathcal{C} = \{U(m,j)\}_{m=1,j=1}^{\mathsf{M},\mathsf{J}}$ be a random codebook whose elements are drawn independently from q_U. Here J is introducing redundancy but since it will be decoded at the receiver we can view it as a dummy message.

Encoding: Instead of using conventional random covering, we use an SMC which acts as follows. Given m and s, the SMC chooses an index j with the probability

$$P_{\mathsf{Enc}}(j|m,s) = \frac{2^{\imath(s;U(m,j))}}{\sum_{\tilde{j}} 2^{\imath(s;U(m,\tilde{j}))}}.$$

Then the encoder transmits $x(U(m,j),s)$ through the channel. Observe that the above SMC resembles a joint-typical encoder of the asymptotic regime. Given m and s, the higher the information between $U(m,j)$ and s, the more likely we choose it at the encoder.

Decoding: In contrast to the point-to-point problem, computing the error probability of SLC is challenging. An SLC uses the induced $P_{M,J|Y}$ by the code. Instead, we use an SMC with the following rule for decoding. Observing y, decoder uses the following SMC to find both the message m and the dummy message j:

$$P_{\mathsf{Dec}}(\hat{m},\hat{j}|y) = \frac{2^{\imath(y;U(\hat{m},\hat{j}))}}{\sum_{\bar{m},\bar{j}} 2^{\imath(y;U(\bar{m},\bar{j}))}}.$$

Analysis: We declare an error if $(\hat{m},\hat{j}) \neq (m,j)$. Observe that the joint distribution of random variables factors as,

$$P_{MJSY\hat{M}\hat{J}}(m,j,s,y,\hat{m},\hat{j}) = \frac{1}{\mathsf{M}} q(s) P_{\mathsf{Enc}}(j|m,s) q(y|U(m,j),s) P_{\mathsf{Dec}}(\hat{m},\hat{j}|y),$$

where $q(y|U(m,j),s) = q_{Y|X,S}\big(y|x(U(m,j),s),s\big)$. The probability of correct decoding is $\mathsf{P}[C] = \sum_{m,j,s,y} P_{MJSY\hat{M}\hat{J}}(m,j,s,y,m,j)$; hence we have:

$$\mathbb{E}\mathsf{P}[C] = \mathbb{E} \sum_{m,j,s,y} \frac{1}{\mathsf{M}} q(s) \frac{2^{\imath(s;U(m,j))}}{\sum_{\tilde{j}} 2^{\imath(s;U(m,\tilde{j}))}} q(y|U(m,j),s) \times \frac{2^{\imath(y;U(m,j))}}{\sum_{\bar{m},\bar{j}} 2^{\imath(y;U(\bar{m},\bar{j}))}} \tag{12}$$

$$= \mathbb{E} \sum_{s,y} \mathsf{J} q(s) \frac{2^{\imath(s;U(1,1))}}{\sum_{\tilde{j}} 2^{\imath(s;U(1,\tilde{j}))}} q(y|U(1,1),s) \times \frac{2^{\imath(y;U(1,1))}}{\sum_{\bar{m},\bar{j}} 2^{\imath(y;U(\bar{m},\bar{j}))}} \tag{13}$$

$$\geq \sum_{s,y} \mathbb{E}_{U(1,1)} \left(\frac{\mathsf{J} q(s) 2^{\imath(s;U(1,1))}}{\mathbb{E}_{\mathcal{C}|U(1,1)} \sum_{\tilde{j}} 2^{\imath(s;U(1,\tilde{j}))}} q(y|U(1,1),s) \frac{2^{\imath(y;U(1,1))}}{\mathbb{E}_{\mathcal{C}|U(1,1)} \sum_{\bar{m},\bar{j}} 2^{\imath(y;U(\bar{m},\bar{j}))}} \right) \tag{14}$$

$$\geq \sum_{s,y} \mathbb{E}_{U(1,1)} \left(\frac{\mathsf{J} q(s) 2^{\imath(s;U(1,1))}}{2^{\imath(s;U(1,1))} + \mathsf{J}} q(y|U(1,1),s) \frac{2^{\imath(y;U(1,1))}}{2^{\imath(y;U(1,1))} + \mathsf{MJ}} \right) \tag{15}$$

$$= \sum_{u,s,y} q(u,s,y) \frac{\mathsf{J}}{2^{\imath(s;u)} + \mathsf{J}} \cdot \frac{2^{\imath(y;u)}}{2^{\imath(y;u)} + \mathsf{MJ}} \tag{16}$$

$$= \mathbb{E}_{USY} \frac{1}{(1 + \mathsf{J}^{-1} 2^{\imath(U;S)})(1 + \mathsf{MJ} 2^{-\imath(U;Y)})}, \tag{17}$$

where (13) is due to symmetry, the main step (14) follows from Jensen inequality for the two-valued convex function $f(x_1,x_2) = \frac{1}{x_1 x_2}$ on $\mathbb{R}_+^2$, (15) follows from the fact that $U(i,j)$ is independent of $U(1,1)$ for $(i,j) \neq (1,1)$ and generated according to q_U, and (16) follows from the fact that $U(1,1)$ is generated according to q_U. ∎

B. Second Order achievability of Gelfand-Pinsker channel

Theorem 3: [3] Given a memoryless state-dependent channel $(q_S, q_{Y|X,S})$ with state known non-causally at the encoder, for any $(q_{U|S}, x(u,s))$, the following rate is (n,ϵ)-achievable:

$$R = I(U;Y) - I(U;S) - \frac{1}{\sqrt{n}} R_D - O\left(\frac{\log n}{n}\right) \tag{18}$$

where

$$R_D = \min_{R:\exists \tilde{R},\ \text{s.t. } [\tilde{R}, R-\tilde{R}]^{\mathsf{T}} \in \mathcal{Q}^{-1}(\mathbb{V}_{\mathsf{GP}},\epsilon)} R, \tag{19}$$

and

$$\mathbb{V}_{\mathsf{GP}} = \mathsf{Cov}([\imath(U;S) \quad -\imath(U;Y)]^{\mathsf{T}}). \tag{20}$$

Proof: We apply (11) to n use of the channel. Assume that $q_{U^n}(u^n) = \prod_{i=1}^n q_U(u_i)$, so (U^n, S^n, Y^n) are i.i.d.. Substituting $\gamma = \frac{1}{2}\log n$ in (11) implies:

$$1 - \epsilon - O(\frac{1}{\sqrt{n}}) = \mathsf{P}\left(\log \mathsf{J} - \imath(U^n;S^n) > \frac{1}{2}\log n, \ \imath(U;Y) - \log \mathsf{MJ} > \frac{1}{2}\log n \right). \tag{21}$$

Let $\log \mathsf{J} = nI(U;S) + \sqrt{n}\tilde{R} + \frac{1}{2}\log n$ and $\log \mathsf{M} = n(I(U;Y) - I(U;S)) - \sqrt{n}R - \log n$. The random variables $\imath(U^n;S^n)$ and $\imath(U^n;Y^n)$ are sum of i.i.d. random variabels.

[3] Concurrently, the same result has been obtained by Watanabe, et. al. [9].

Applying multi-dimensional berry-esseen CLT [10] to (21) implies that

$$1-\epsilon-O(\frac{1}{\sqrt{n}})=\mathsf{P}_G\left(\left[\begin{array}{c}G_1\\G_2\end{array}\right]\leq\left[\begin{array}{c}\tilde{R}\\R-\tilde{R}\end{array}\right]\right),\quad(22)$$

where $G=[G_1\ G_2]^{\mathsf{T}}$ is a multidimensional normal r.v. with zero mean and $\mathsf{Cov}G=\mathbb{V}_{\mathsf{GP}}$. Using the definition of $\mathcal{Q}^{-1}(\mathsf{V},\epsilon)$ and smoothness of distribution of normal r.v., we get

$$[\tilde{R},R-\tilde{R}]^{\mathsf{T}}\in\mathcal{Q}^{-1}(\mathbb{V}_{\mathsf{GP}},\epsilon)+O\left(\frac{\log n}{\sqrt{n}}\right),\quad(23)$$

which completes the proof. ∎

V. Broadcast channel

Consider the problem of transmission of private messages over a broadcast channel. Let $q_{Y_1Y_2|X}$ be channel transition probability. We prove a one-shot version of Marton with two auxiliaries. A similar theorem is proved for Marton with common message and involving auxiliary rv U_0 in [11].

Theorem 4: Given any q_{U_1,U_2} and function $x(u_1,u_2)$, there is a code for a single use of the channel whose probability of correct decoding is bounded from below by

$$\mathbb{E}\left[\left(1+(\mathsf{J}_1\mathsf{J}_2)^{-1}2^{\imath(U_1;U_2)}\right)\prod_{k=1}^{2}(1+\mathsf{M}_k\mathsf{J}_k2^{-\imath(U_k;Y_k)})\right]^{-1},$$

where $\mathsf{J}_1,\mathsf{J}_2>0$ are arbitrary integers. Moreover, loosening this bound gives the following upper bound on error probability of the code,

$$\begin{aligned}\mathsf{P}\big[&\log\mathsf{J}_1\mathsf{J}_2-\imath(U_1;U_2)<\gamma\ ,\quad\text{or}\\&\imath(U_1;Y_1)-\log\mathsf{M}_1\mathsf{J}_1<\gamma\ ,\ \text{or}\\&\imath(U_2;Y_2)-\log\mathsf{M}_2\mathsf{J}_2<\gamma\big]+7\times2^{-\gamma},\end{aligned}\quad(24)$$

where γ is any positive number.

Remark 3: Verdu derives a one-shot bound for the same problem in [8]. He derives the bound by proposing a one-shot covering and packing lemma. However to get access to the boundary of Marton's inner bound one needs a mutual covering lemma (since time sharing is not possible in one-shot and not useful in finite block length regime). For this reason Verdu's result seems to be weaker than ours. Our technique allows us to bypass the need for developing a one-shot version of the mutual covering lemma.

Proof: We only prove the lower bound on probability of correct decoding. Derivation of the loosened bound can be found in [11]. Let $\mathcal{C}=\mathcal{C}_1\times\mathcal{C}_2=\{U_1(m_1,j_1)\}_{m_1=1,j_1=1}^{\mathsf{M}_1,\mathsf{J}_1}\times\{U_2(m_2,j_2)\}_{m_2=1,j_2=1}^{\mathsf{M}_2,\mathsf{J}_2}$ be a random product codebook, in which the codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ are generated independently and the elements of the codebook $\mathcal{C}_k,k=1,2$ s are drawn independently from q_{U_1}. Thus the codebook is generated according to $r_{U_1U_2}=q_{U_1}q_{U_2}\neq q_{U_1U_2}$. Here J_1,J_2 are introducing redundancy but since it will be decoded at the receiver we can view these as dummy messages.

Encoding: Instead of using conventional mutual covering, we use an SMC which acts as follows. Given m_1,m_2, the SMC chooses indices j_1,j_2 with the probability

$$P_{\mathsf{Enc}}(j_1,j_2|m_1,m_2)=\frac{2^{\imath_q(U_1(m_1,j_1);U_2(m_2,j_2))}}{\sum_{\tilde{j}_1,\tilde{j}_2}2^{\imath_q(U_1(m_1,\tilde{j}_1);U_2(m_2,\tilde{j}_2))}}.\quad(25)$$

Then the encoder transmits $x(U_1(m_1,j_1),U_2(m_2,j_2))$ through the channel. Observe that we generate codewords according to $r_{U_1U_2}$ but compute the informations $\imath_q$ using $q_{U_1U_2}$. This resembles the Marton coding scheme where we generate U_1^n and U_2^n independently but choose the jointly typical ones for transmission.

Decoding: We again use an SMC for decoding. Observing y_k, decoder k uses the following SMC to find both the message m_k and the dummy message j_k:

$$P_{\mathsf{Dec}_k}(\hat{m}_k,\hat{j}_k|y)=\frac{2^{\imath_q(y_k;U_k(\hat{m}_k,\hat{j}_k))}}{\sum_{\bar{m}_k,\bar{j}_k}2^{\imath_q(y_k;U_k(\bar{m}_k,\bar{j}_k))}}.$$

Analysis: Observe that the joint distribution of random variables factors as,

$$\begin{aligned}P(m_{1:2},j_{1:2},y_{1:2},\hat{m}_{1:2},\hat{j}_{1:2})=&\frac{1}{\mathsf{M}_1\mathsf{M}_2}P_{\mathsf{Enc}}(j_{1:2}|m_{1:2})\\&q\big(y_{1:2}|U_1(m_1,j_1),U_2(m_2,j_2)\big)\prod_{k=1}^{2}P_{\mathsf{Dec}_k}(\hat{m}_k,\hat{j}_k|y_k),\end{aligned}$$

where $q\big(y_{1:2}|U_1(m_1,j_1),U_2(m_2,j_2)\big)$ is equal to $q_{Y_{1:2}|X}\big(y_{1:2}|x(U_1(m_1,j_1),U_2(m_2,j_2))\big)$. We make an error if either of the decoders fail. We analyse the error probability for both of the decoders without using a union bound in [11]. However to save some space here, we only compute the probability of correct decoding at decoder one. The probability of correct decoding can be bounded from below by $\mathsf{P}[C]\geq\sum_{m_{1:2},j_{1:2},y_1}P(m_{1:2},j_{1:2},y_1,\hat{M}_1=m_1,\hat{J}_1=j_1)$ (where we omitted $\hat{m}_2,\hat{j}_2$ from the probability), hence skipping similar symmetry arguments we have:

$$\begin{aligned}\mathbb{E}\mathsf{P}[C]\geq&\ \mathbb{E}\sum_{y_1}\mathsf{J}_1\mathsf{J}_2\frac{2^{\imath_q(U_1(1,1);U_2(1,1))}}{\sum_{\tilde{j}_1,\tilde{j}_2}2^{\imath_q(U_1(1,\tilde{j}_1);U_2(1,\tilde{j}_2))}}\\&q(y_1|U_1(1,1),U_2(1,1))\frac{2^{\imath_q(y_1;U_1(1,1))}}{\sum_{\bar{m}_1,\bar{j}_1}2^{\imath_q(y_1;U_1(\bar{m}_1,\bar{j}_1))}}&(26)\\\geq&\sum_{y_1}\mathbb{E}_{U_{1:2}(1,1)}\left(\frac{\mathsf{J}_1\mathsf{J}_2\,2^{\imath_q(U_1(1,1);U_2(1,1))}}{\mathbb{E}_{\mathcal{C}|U_{1:2}(1,1)}\sum_{\tilde{j}_1,\tilde{j}_2}2^{\imath_q(U_1(1,\tilde{j}_1);U_2(1,\tilde{j}_2))}}\right.\\&\left.q(y_1|U_{1:2}(1,1))\frac{2^{\imath_q(y_1;U_1(1,1))}}{\mathbb{E}_{\mathcal{C}|U_{1:2}(1,1)}\sum_{\bar{m},\bar{j}}2^{\imath_q(y_1;U_1(\bar{m},\bar{j}))}}\right)&(27)\\\geq&\sum_{y_1}\mathbb{E}_{U_{1:2}(1,1)}\left(\frac{\mathsf{J}_1\mathsf{J}_2\,2^{\imath_q(U_1(1,1);U_2(1,1))}}{2^{\imath_q(U_1(1,1);U_2(1,1))}+\mathsf{J}_1\mathsf{J}_2}\right.\\&\left.q(y_1|U_{1:2}(1,1))\frac{2^{\imath_q(y_1;U_1(1,1))}}{2^{\imath_q(y_1;U_1(1,1))}+\mathsf{M}_1\mathsf{J}_1}\right)&(28)\\=&\sum_{u_1,u_2,y_1}q(u_1,u_2,y_1)\frac{\mathsf{J}_1\mathsf{J}_2}{2^{\imath_q(u_1;u_2)}+\mathsf{J}_1\mathsf{J}_2}\cdot\frac{2^{\imath_q(y_1;u_1)}}{2^{\imath_q(y_1;u_1)}+\mathsf{M}_1\mathsf{J}_1}&(29)\\=&\mathbb{E}_q\frac{1}{(1+(\mathsf{J}_1\mathsf{J}_2)^{-1}2^{\imath_q(U_1;U_2)})(1+\mathsf{M}_1\mathsf{J}_1\mathsf{J}_2^{-\imath_q(U_1;Y_1)})}\end{aligned}$$

where (26) is due to symmetry, the main step (27) follows from Jensen inequality for the two-valued convex function

$f(x_1,x_2)=\dfrac{1}{x_1x_2}$ on $\mathbb{R}_+^2$ (for the case of probability of correct decoding at both decoders discussed in [11], we use the three-valued convex function $f(x_1,x_2,x_3)=\dfrac{1}{x_1x_2x_3}$). The expectation in (27) is over $U_{1:2}(1,1)$ of the codebook generation distributed according to $r_{U_1U_2}$. Equation (28) follows from the following equations:

$$\mathbb{E}_{\mathcal{C}|U_{1:2}(1,1)}2^{\imath_q(U_1(1,1);U_2(1,\tilde{j}_2))}=\sum_{u_2}q(u_2)2^{\imath_q(U_1(1,1);u_2)}$$
$$=\sum_{u_2}q(u_2|U(1,1))=1,\ \tilde{j}_2\neq 1, \tag{30}$$
$$\mathbb{E}_{\mathcal{C}|U_{1:2}(1,1)}2^{\imath_q(U_1(1,\tilde{j}_1);U_2(1,1))}=1,\ \tilde{j}_1\neq 1, \tag{31}$$
$$\mathbb{E}_{\mathcal{C}|U_{1:2}(1,1)}2^{\imath_q(U_1(1,\tilde{j}_1);U_2(1,\tilde{j}_2))}=1,\ \tilde{j}_1\neq 1\ \tilde{j}_2\neq 1, \tag{32}$$

where in (30) we use the fact that $U_2(1,\tilde{j}_2)$ is independent of $(U_1(1,1),U_2(1,1))$ for $\tilde{j}_2\neq 1$ and generated according to q_{U_2}. (31) and (32) follows similarly. Finally and (29) follows from the fact that $(U_1(1,1),U_2(1,1))$ is generated according to product distribution $r_{U_1U_2}=q_{U_1}q_{U_2}$. ■

VI. Berger-Tung

Consider the problem of distributed lossy compression. Let $q_{S_1S_2}$ be the joint distribution of the sources and $\mathsf{d}_k(s_k,\hat{s}_k), k=1,2$, be two distortion measures. We will use the *probability of excess distortion* as the criterion for measuring the reliability of the system.

Definition 3: An $(\mathsf{M}_1,\mathsf{M}_2)$-code for distributed lossy compression consists of (possibly stochastic) encoders $\varphi_k:\mathcal{S}_k\mapsto[1:\mathsf{M}_k], k=1,2$, and (possibly stochastic) decoder $\psi:[1:\mathsf{M}_1]\times[1:\mathsf{M}_2]\mapsto\widehat{S}_1\times\widehat{S}_2$. Given a pair of distortion levels $(\mathsf{D}_1,\mathsf{D}_2)$, the probability of excess distortion is defined by,

$$\mathsf{P}[\mathcal{E};\mathsf{D}_1,\mathsf{D}_2]:=\mathsf{P}[\mathsf{d}_1(S_1,\widehat{S}_1)>\mathsf{D}_1\ \vee\ \mathsf{d}_2(S_2,\widehat{S}_2)>\mathsf{D}_2].$$

Also, we define the probability of *no-excess distortion* by

$$\mathsf{P}[C]:=\mathsf{P}[\mathsf{d}_1(S_1,\widehat{S}_1)\leq\mathsf{D}_1,\ \mathsf{d}_2(S_2,\widehat{S}_2)\leq\mathsf{D}_2].$$

We prove a one-shot version of the result of Berger-Tung for this problem.

Theorem 5: Given any pmf $q_{U_1|S_1}q_{U_2|S_2}$ and functions $\hat{s}_k(u_1,u_2), k=1,2$, there is an $(\mathsf{M}_1,\mathsf{M}_2)$-code for a single use of the sources whose probability of *no-excess distortion* is bounded from below by

$$\mathsf{P}[C]\geq\mathbb{E}\frac{\mathbf{1}\left\{\mathsf{d}_k\left(S_k,\hat{s}_k(U_1,U_2)\right)\leq\mathsf{D}_k,\ k=1,2\right\}}{\left(1+\mathsf{J}_1\mathsf{J}_2 2^{-\imath(U_1;U_2)}\right)\prod_{k=1}^2(1+(\mathsf{M}_k\mathsf{J}_k)^{-1}2^{\imath(S_k;U_k)})},$$

where J_1 and J_2 are arbitrary integers.

Remark 4: The term $1+(\mathsf{M}_k\mathsf{J}_k)^{-1}2^{\imath(S_k;U_k)},\ \ k=1,2$ corresponds to a covering of S_k with U_k in the asymptotic regime. The term $\mathsf{J}_1\mathsf{J}_2 2^{-\imath(U_1;U_2)}$ corresponds to a mutual packing lemma in the asymptotic regime.

Proof:

Random codebook generation: Let $\mathcal{C}=\mathcal{C}_1\times\mathcal{C}_2=\{U_1(m_1,j_1)\}_{m_1=1,j_1=1}^{\mathsf{M}_1,\mathsf{J}_1}\times\{U_2(m_2,j_2)\}_{m_2=1,j_2=1}^{\mathsf{M}_2,\mathsf{J}_2}$ be a random product codebook, in which the codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ are generated independently and the elements of the codebook $\mathcal{C}_k, k=1,2$ are drawn independently from q_{U_k}. Thus the codebook is generated according to $r_{U_1U_2}=q_{U_1}q_{U_2}\neq q_{U_1U_2}$.

Encoding: Encoder $k=1,2$ uses an SMC followed by a random binning to obtain the pair (m_k,j_k). First given s_k, the SMC chooses a pair (m_k,j_k) with the probability

$$P_{\mathsf{Enc}_k}(m_k,j_k|s_k)=\frac{2^{\imath_q(s_k;U(m_k,j_k))}}{\sum_{\tilde{m}_k,\tilde{j}_k}2^{\imath_q(s_k;U(\tilde{m}_k,\tilde{j}_k))}}.$$

Then the encoder transmits m_k to the decoder.

Decoding: We use a SMC for decoding w.r.t. the received indices (m_1,m_2). Observing (m_1,m_2), decoder uses the following SMC to find both the j_1 and j_2 and thus (U_1,U_2):

$$P_{\mathsf{Dec}}(\hat{j}_1,\hat{j}_2|m_1,m_2)=\frac{2^{\imath_q(U_1(m_1,\hat{j}_1);U_2(m_2,\hat{j}_2))}}{\sum_{\bar{j}_1,\bar{j}_2}2^{\imath_q(U_1(m_1,\bar{j}_1);U_2(m_2\bar{j}_2))}}.$$

Remark: Observe that the above SMC resembles a joint-typical decoder for the mutual packing lemma in the asymptotic regime. It can be considered as a dual of the SMC encoder of equation (25) that corresponded to a mutual covering lemma in the asymptotic regime.

Analysis: We consider two error events:

1) The decoder fails to recover the correct pair (j_1,j_2), i.e. $(\hat{j}_1,\hat{j}_2)\neq(j_1,j_2)$.
2) One of the distortions corresponding to the pair (j_1,j_2) exceeds the designated levels, i.e. $\mathsf{d}_k\big(s_k,\hat{s}_k(U_1(m_1,j_1),U_2(m_2,j_2))\big)>\mathsf{D}_k$ for some $k\in\{1,2\}$.

The analysis continues in the same way as in the proof of broadcast channel and can be found in [11]. ■

References

[1] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informations theorie," in Trans. Third. Prague Conf. Inf. Theory, 1962, pp. 689723.
[2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, 56(5), 2307 2359, 2010.
[3] V. Kostina, S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime", arXiv:1209.1317, Sep. 2012.
[4] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in Allerton Conference, 2011, arXiv:1109.6310.
[5] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," arXiv:1201.3901, Feb 2012.
[6] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing ", Physical Review Letters, 2012.
[7] M. Berta, M. Christandl and R. Renner, "The quantum reverse shannon theorem based on one-shot information theory," Commun. Math. Phys. 306, 579615, 2011.
[8] S. Verdú, "Non-Asymptotic Achievability Bounds in Multiuser Information Theory", Allerton Conference, Oct. 2012.
[9] S. Watanabe, S. Kuzuoka, V. Y. F. Tan, "Non-asymptotic and second-order achievability bounds for coding with side-information," arXiv:1301.6467.
[10] V. Bentkus, "On the dependence of the Berry–Esseen bound on dimension," Journal of Statistical Planning and Inference, 113(2), 385-402, 2003.
[11] M. H. Yassaee, M. R. Aref and A. Gohari, "A technique for deriving one-shot achievability results in network information theory", arXiv:1303.0696.