Non-Asymptotic Output Statistics of Random Binning and Its Applications

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*—In this paper we develop a finite blocklength version of the Output Statistics of Random Binning (OSRB) framework. This framework is shown to be optimal in the point-to-point case. New second order regions for broadcast channel and wiretap channel with strong secrecy criterion are derived.**

I. Introduction

Output Statistics of Random Binning (OSRB) is a new framework for proving achievability results [1]. It works by converting channel coding problems into source coding problems, and uses the induced pmf of the source coding side to design encoders for the channel coding side. The goal is to make the total variation distance between the joint pmf of the source coding side and channel coding side close to zero so that all the performance analysis can be dealt with at the source coding side where Slepian-Wolf (S-W) theorem can be invoked. Thus the OSRB technique is not based on the usual covering and packing lemmas.

Originally studied by Strassen [2], there has been a recent surge of works on *finite blocklength information theory* following the work of Polyanskiy et al [3] (see for instance [4]-[7]). In this paper we develop a finite blocklength version of the OSRB framework. We show that this method is optimal in the point-to-point channel and can directly give us the channel dispersion $\mathbb{E}[\mathsf{Var}(\imath(X;Y))|X]$.[1] We also use the technique to derive the second order region for broadcast channel (that recovers Marton's inner bound in the asymptotic case) and for wiretap channel with strong secrecy criterion (that improves the result of [8] for second order coding rate). Scenarios such as broadcast wiretap channel can be also dealt with using this technique but have been left out for a more complete version of this draft.

OSRB is based on two theorems: the S-W theorem and another theorem that may be considered as its dual. To develop a finite blocklength version of the OSRB, we first find a one-shot version of these two main theorems. By one-shot we mean that only a single use of the resource is allowed. To get finite blocklength results, we then apply this result to a product of n use of the network. The resulting dispersion at the output can be either due to the dispersion in the input code or to the inherent dispersion of the channel. To avoid the input dispersion, we use a uniform distribution over a fixed type in the source coding side of the problem. In this sense this differs from the original asymptotic OSRB where we use a completely i.i.d. distribution in the source coding side of the problem.

This paper is organized as follows: some definitions and notations are given in Section II. One-shot version of the two main theorems of the OSRB are given in Section III. We then apply the technique to a couple of problems in Section IV. To illustrate the use of the technique we begin by recovering the known result on dispersion for the point-to-point channel in Subsection IV-A. In Subsections IV-B and IV-C, we apply the technique to broadcast channel and wiretap channel.

II. Definitions

Definition 1: Given a pmf $p_{X,Y}$, the conditional information of x given y is defined by

$$h_p(x|y) := \log \frac{1}{p_{X|Y}(x|y)}.$$

Also, the *information density* $\imath_p(x;y)$ is defined by

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

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

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

Notation: In this paper, we use $X_{\mathcal{V}}$ to denote $(X_v : v \in \mathcal{V})$ and $p^U_{\mathcal{A}}$ to denote the uniform distribution over the set $\mathcal{A}$. The total variation between two pmf's p and q on the same alphabet $\mathcal{X}$, is defined by $\|p(x)-q(x)\|_1 := \frac{1}{2}\sum_x |p(x)-q(x)|$.

III. One-shot Output Statistics of Random Binning

Let $(X_{\mathcal{V}}, Z)$ be a set of discrete sources distributed according to a joint pmf $p_{X_{\mathcal{V}},Z}$ on a finite set $(\prod_{v\in\mathcal{V}} \mathcal{X}_v) \times \mathcal{Z}$. A distributed random binning consists of a set of random mappings $\mathcal{B}_v : \mathcal{X}_v \to [1 : \mathsf{M}_v]$, $v \in \mathcal{V}$, in which $\mathcal{B}_v$ maps each sequence of $\mathcal{X}_v$ uniformly and independently to the set $[1 : \mathsf{M}_v]$. We use B_v as a shorthand for rv $\mathcal{B}_v(X_v)$, and $B_{\mathcal{V}}$ or $B_{\mathcal{V}}(X_{\mathcal{V}})$ as a shorthand for rv $(\mathcal{B}_v(X_v))_{v\in\mathcal{V}}$. A distributed

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

[1] Direct proofs for this formula have also been obtained by Wang et. al. using a different technique [6].

random binning induces the following *random pmf*[2] on the set $\mathcal{X}_{\mathcal{V}}^n \times \mathcal{Z}^n \times \prod_{v\in\mathcal{V}}[1:\mathsf{M}_v]$,

$$P(x_{\mathcal{V}}, z, b_{\mathcal{V}}) = p(x_{\mathcal{V}}, z)\prod_{v\in\mathcal{V}}\mathbf{1}\{\mathcal{B}_v(x_v) = b_v\}. \qquad (1)$$

The asymptotic OSRB introduced in [1] relies on the S-W theorem as well as Theorem 1 of [1] that implies independence of random bin indices under certain conditions. To set up a *non-asymptotic* framework, we generalize the S-W theorem and Theorem 1 of [1] to the case of a single channel use. Let us begin with the latter:

Theorem 1 (One-shot OSRB): Given $p_{X_{\mathcal{V}},Z}$, for any pmf t_Z and any positive real γ, the random pmf of eq. (1) satisfies

$$\mathbb{E}\left\|P(b_{\mathcal{V}}, z) - p^U(b_{\mathcal{V}})p(z)\right\|_1 \leq p_{X_{\mathcal{V}}Z}(\mathcal{S}_\gamma(p\|t)^c) + 2^{\frac{|\mathcal{V}|-\gamma}{2}-1},$$

where the expectation is over the randomness of binning and the set $\mathcal{S}_\gamma(p\|t) \subset \mathcal{X}_{\mathcal{V}}^n \times \mathcal{Z}^n$ is defined as follows:

$$\mathcal{S}_\gamma(p\|t) := \{(x_{\mathcal{V}}, z) : \forall\emptyset \neq \mathcal{S} \subseteq\mathcal{V}, h_p(x_{\mathcal{S}}, z) - h_t(z) - \sum_{v\in\mathcal{S}}\log\mathsf{M}_v > \gamma\}.$$

Remark 1: This theorem implies [1, Theorem 1] by evaluating it for the product $p_{X_{\mathcal{V}}^n,Z^n} = \prod_{i=1}^n p_{X_{\mathcal{V},i},Z_i}$. Set $t_{Z^n} = p_{Z^n}$ and $\gamma = n\delta$ for a sufficiently small value of $\delta > 0$ that we discuss later. Then the term $2^{\frac{|\mathcal{V}|-1-\gamma}{2}}$ converges to zero. The first term converges to zero as well if $\mathcal{S}_\gamma(p\|t)$ includes almost all of the typical set $\mathcal{T}_p$. For any jointly typical $(x_{\mathcal{V}}^n, z^n)$, the terms $h_p(x_{\mathcal{S}}^n, z^n)$ and $h_p(z^n)$ are close to $nH_p(X_{\mathcal{S}}, Z)$ and $nH_p(Z)$, respectively. Thus as long as $H_p(X_{\mathcal{S}}|Z) > \frac{1}{n}\sum_{v\in\mathcal{S}}\log M_v$ we can choose $\gamma = n\delta$ for a $\delta > 0$ such that the inequalities defining $\mathcal{S}_\gamma(p\|t)$ holds.

Remark 2: The rv Z in the statement of the above theorem is of use in problems with secrecy constraints.

Proof: See [10]. ■

One shot S-W coding: Here we want to bound the error probability of decoding a single copy of the source $X_{\mathcal{V}}$ when the decoder has access to the side information Z as well as the bin indices $B_{\mathcal{V}}$. An optimal decoder uses ML decoding. However we use an *stochastic* variation of MAP for the decoding with a more tractable analysis. The decoder draws $\hat{x}_{\mathcal{V}}$ from the conditional pmf $P_{X_{\mathcal{V}}|Z,B_{\mathcal{V}}}(\hat{x}_{\mathcal{V}}|y, b_{\mathcal{V}})$, where P is the induced probability by the random binning. More specifically

$$P_{X_{\mathcal{V}}|Z,B_{\mathcal{V}}}(\hat{x}_{\mathcal{V}}|z, b_{\mathcal{V}}) = \frac{p(\hat{x}_{\mathcal{V}}|z)\mathbf{1}(B_{\mathcal{V}}(\hat{x}_{\mathcal{V}}) = b_{\mathcal{V}})}{\sum_{\bar{x}_{\mathcal{V}}} p(\bar{x}_{\mathcal{V}}|z)\mathbf{1}(B_{\mathcal{V}}(\bar{x}_{\mathcal{V}}) = b_{\mathcal{V}})}.$$

We refer this decoder as a *stochastic likelihood coder* (SLC). See [11] for a motivation of SLC and the justification for using a stochastic decoder. For some technical reasons,we can more generally use a *mismatch* SLC corresponding to an arbitrary pmf $t_{X_{\mathcal{V}},Z}$ instead of p in the above expression,[3] that is,

$$T_{X_{\mathcal{V}}|Z,B_{\mathcal{V}}}(\hat{x}_{\mathcal{V}}|z, b_{\mathcal{V}}) = \frac{t(\hat{x}_{\mathcal{V}}|z)\mathbf{1}(B_{\mathcal{V}}(\hat{x}_{\mathcal{V}}) = b_{\mathcal{V}})}{\sum_{\bar{x}_{\mathcal{V}}} t(\bar{x}_{\mathcal{V}}|z)\mathbf{1}(B_{\mathcal{V}}(\bar{x}_{\mathcal{V}}) = b_{\mathcal{V}})}.$$

Roughly speaking, the reason for introducing a mismatch SLC is that we will need to work with input codewords of the same type to reduce the total dispersion, rather than with codewords generated from an i.i.d. distribution. However we need independence to be able to use the Berry-Esseen CLT at a later stage. A mismatch SLC allows us to simultaneously employ an independent and a non-independent distribution.

Theorem 2 (One-shot S-W): Given $p_{X_{\mathcal{V}},Z}$ and any pmf $t_{X_{\mathcal{V}},Z}$, the expected value of the probability of correct decoding of a mismatch SLC associated with t is bounded from below by

$$\mathbb{E}\mathsf{P}[C] \geq \mathbb{E}_p\frac{1}{1 + \sum_{\emptyset\neq\mathcal{S}\subseteq\mathcal{V}}\mathsf{M}_{\mathcal{S}}^{-1}2^{h_t(X_{\mathcal{S}}|X_{\mathcal{S}}^c,Z)}}, \qquad (2)$$

where $\mathsf{M}_{\mathcal{S}} = \prod_{v\in\mathcal{S}}\mathsf{M}_v$. Moreover, this bound can be weakened to give the following bound on the error probability of mismatch SLC,

$$\mathbb{E}\mathsf{P}[\mathcal{E}] \leq p_{X_{\mathcal{V}}Z}(\mathcal{S}_\gamma(t_{X_{\mathcal{V}},Z})^c) + (2^{|\mathcal{V}|} - 1)2^{-\gamma}, \qquad (3)$$

where γ is an arbitrary positive number and

$$\mathcal{S}_\gamma(t_{X_{\mathcal{V}},Z}) := \{(x_{\mathcal{V}}, z) : \forall\emptyset \neq \mathcal{S} \subseteq \mathcal{V}, \sum_{v\in\mathcal{S}}\log\mathsf{M}_v - h_t(x_{\mathcal{S}}|z) > \gamma\}. \qquad (4)$$

Remark 3: Using this theorem one can derive finite blocklength analogs of the S-W theorem for i.i.d. or non-i.i.d. sources. Since we choose the codewords from a fixed type, we use this theorem in its non-i.i.d. form. I.i.d. forms of the S-W theorem have been previously obtained by [7].

Proof: We only prove the inequality (2) for the special case of $|\mathcal{V}| = 1$. For the complete proof, see [10]. The probability of correct decoding can be written as,

$$\mathsf{P}[C] = \sum_{x,b,z} p(x, z)\mathbf{1}(B(x) = b)T_{X|Z,B}(x|z, b).$$

We have,

$$\mathbb{E}\mathsf{P}[C] = \mathbb{E}\sum_{x,z,b} p(x, z)\mathbf{1}(B(x) = b)\frac{t(x|z)}{\sum_{\bar{x}} t(\bar{x}|z)\mathbf{1}(B(\bar{x}) = b)} \qquad (5)$$

$$= \mathsf{M}\mathbb{E}\sum_{x,z} p(x, z)\mathbf{1}(B(x) = 1)\frac{t(x|z)}{\sum_{\bar{x}} t(\bar{x}|z)\mathbf{1}(B(\bar{x}) = 1)} \qquad (6)$$

$$= \mathsf{M}\sum_{x,z}\mathbb{E}_{B(x)}\mathbb{E}_{\{B(\bar{x}),\bar{x}\neq x\}}p(x, z)\frac{t(x|z)\mathbf{1}(B(x) = 1)}{\sum_{\bar{x}} t(\bar{x}|z)\mathbf{1}(B(\bar{x}) = 1)} \qquad (7)$$

$$\geq \mathsf{M}\sum_{x,z}\mathbb{E}_{B(x)}p(x, z)\frac{t(x|z)\mathbf{1}(B(x) = 1)}{\mathbb{E}_{\{B(\bar{x}),\bar{x}\neq x\}}\sum_{\bar{x}} t(\bar{x}|z)\mathbf{1}(B(\bar{x}) = 1)} \qquad (8)$$

$$= \mathsf{M}\sum_{x,z}\mathbb{E}_{B(x)}p(x, z)\frac{t(x|z)\mathbf{1}(B(x) = 1)}{t(x|z)\mathbf{1}(B(x) = 1) + \mathsf{M}^{-1}(1 - t(x|z))} \qquad (9)$$

$$\geq \mathsf{M}\sum_{x,z}\mathbb{E}_{B(x)}p(x, z)\frac{t(x|z)\mathbf{1}(B(x) = 1)}{t(x|z)\mathbf{1}(B(x) = 1) + \mathsf{M}^{-1}} \qquad (10)$$

$$= \sum_{x,z} p(x, z)\frac{t(x|z)}{t(x|z) + \mathsf{M}^{-1}} = \mathbb{E}_p\frac{1}{1 + \mathsf{M}^{-1}2^{h_t(x|z)}}, \qquad (11)$$

where (6) is due to the symmetry, (8) follows from the Jensen inequality for the convex function $f(x) = \frac{1}{x}$ on the $\mathbb{R}_+$ and

[2] The pmf is random due to the random binning assignment in the protocol.

[3] The pmf $t_{X_{\mathcal{V}},Z}$ should not be confused with the one used in Thm 1 where it is only defined on Z.

(9) follows from the fact that $B(\bar{x})$ and $B(x)$ are independent for any $\bar{x} \neq x$. ■

IV. Applications of non-asymptotic OSRB

To illustrate the use of the tools introdued in the previous section, we recover a finite blocklength result for the point to point channel coding, and prove new results for broadcast channel and wiretap channel. Since the structure of the proofs are similar, we have tried to provide a detailed proof for the simplest case, i.e. the point-to-point channel and outline other proofs have less details. See [10] for the full proofs.

A. Point to point channel coding

Consider a DMC channel $q_{Y|X}$. We will recover the result of [3] that there is an (n, ϵ)-code with rate

$$R(n, \epsilon) = I(X;Y) + \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) - O\left(\frac{\log n}{n}\right), \qquad (12)$$

for any arbitrary input pmf q_X where $V = \mathbb{E}\left[\mathsf{Var}_{q_{Y|X}}(\imath(X;Y)|X)\right]$. Our framework is divided into two steps: in the first step we obtain a one-shot achievable rate following the OSRB technique. In the second step we use Theorem 1 and Theorem 2 for the n uses of the channel, to approximate the achievable rate.

Step 1: One-shot OSRB: Just like the asymptotic OSRB, the first step is itself divided into three parts. In the first part we start from a source coding problem, use random binning and then find an upper bound on the error probability. In the second part, we use the joint pmfs of the source coding side of the problem to design a concrete encoder-decoder for the channel coding with one exception: the encoder-decoder is assisted with a common randomness that does not really exist in the model (to be removed in third part). We will find upper bounds on the total variation distance of the joint induced pmf's between all r.v.'s in the two parts. The bounds on the error probability of S-W coding and the total variation distance of the joint induced pmf's give a bound on the error probability of encoder-decoder of the part two. In the third part, we eliminate the common randomness given to the second protocol without disturbing the probability of error. This makes the designed encoder-decoder in the second part useful for code construction.

Part 1: Source coding problem and random binning: We start from a different problem of source coding; we will use the pmf induced by this problem to construct our channel code in the next part. Let (X, Y) be distributed according to $q(x, y) = q(x)q(y|x)$. We define two random mappings on $\mathcal{X}$ as follows: to each x, we assign two random bin indices $m \in [1 : \mathsf{M}]$ and $f \in [1 : \mathsf{F}]$, uniformly and independently. This induces a joint pmf on M, F, X which we denote by $P_s(m, f, x)$. Suppose that the decoder chooses a $t_{X,Y}$ and uses a mismatched decoder $T(\hat{x}|y, f)$ constructed using $t_{X,Y}$. Then the induced random pmf is $P_s(x, y, m, f, \hat{x}) = q(x, y)P_s(m, f|x)T(\hat{x}|f, y)$. Invoking Theorem 2 with rv Z being a constant, one can derive an upper bound ϵ_{Dec} on the expectation of error probability that only depends on F (and not on M). This upper bound is provided later in equation (17) for the finite blocklength coding.

Part 2: Designing encoder-decoder assisted with a shared randomness: Returning to the channel coding problem we assume that there is a shared randomness F available at both the encoder and decoder, which is independent of the message and uniformly distributed over $[1 : \mathsf{F}]$. This shared randomness does not exist in the original setup and we will eliminate it later. The encoder uses the conditional pmf $P_s(x|m, f)$ of the source coding problem. The decoder uses the mismatched decoder $T(\hat{x}|y, f)$ to find $\hat{x}$ and thereby an estimate of the message $\hat{m}$. The induced random pmf is $P_c(x, y, m, f, \hat{x}) = p^U(m, f)P_s(x|m, f)p(y|x)T(\hat{x}|f, y)$. We have

$$\|P_s(x, y, m, f, \hat{x}) - P_c(x, y, m, f, \hat{x})\|_1 = \left\|P_s(m, f) - p^U(m, f)\right\|_1. \qquad (13)$$

Given M and F, Theorem 1 gives an upper bound ϵ_{Apx} on the expectation of the total variation distance between P_s and P_c. Observe that using P_c instead of P_s changes the probability of error by at most ϵ_{Apx}. Thus the expected error probability $\mathbb{E}_{\mathcal{B}}\mathsf{P}[\mathcal{E}]$ of the channel coding is bounded above by $\epsilon_{\mathsf{Dec}} + \epsilon_{\mathsf{Apx}}$.

Part 3: Eliminating shared randomness: Using the law of iterated expectation, we have $\mathbb{E}_{\mathcal{B}}\mathsf{P}[\mathcal{E}] = \mathbb{E}_{\mathcal{B},F}\mathsf{P}[\mathcal{E}|F] \leq \epsilon_{\mathsf{Dec}} + \epsilon_{\mathsf{Apx}}$. Thus there exists a fixed binning and an instance f^* of F, such that the encoder $p_s(x|m, f^*)$ and the mismatched decoder $T(\hat{x}|y, f^*)$ results in a pair of encoder-decoder with error probability of at most $\epsilon_{\mathsf{Dec}} + \epsilon_{\mathsf{Apx}}$.

Step 2: Non-asymptotic analysis: We would apply the one shot OSRB bound to n i.i.d. repetitions of the DMC $q_{Y|X}$. In [1], we started from an i.i.d. input for the source coding part. Although using an i.i.d. distribution makes evaluation of Theorem 1 and Theorem 2 simple, but this does not yield an optimal strategy. This is due to the fact that an i.i.d. input causes a dispersion in addition to the inherent dispersion of the channel. To avoid input dispersion, we choose channel input sequences with the same type.

Let $\mathsf{M} = 2^{nR}$ and $\mathsf{F} = 2^{n\tilde{R}}$. For a given q_X and n, we can find a n-type $\Phi_X^{(n)}$ such that the infinity norm $\|\Phi_X^{(n)} - q_X\|_\infty \leq \frac{1}{n}$. To prove (12), assume that the p_{X^n} is a uniform distribution over the set $\mathcal{T}_{\Phi_X^{(n)}}$ of sequences with the type $\Phi_X^{(n)}$. The known bounds on the size of typical sets imply that there exists L such that for any n, $\log|\mathcal{T}_{\Phi_X^{(n)}}| \geq nH_{\Phi_X^{(n)}}(X) - L\log n$. Setting $\gamma = \log n$, Z a constant and $|\mathcal{V}| = 1$ in Theorem 1 gives the following bound on the right hand side of equation (13) and thus on ϵ_{Apx}:

$$\epsilon_{\mathsf{Apx}} \leq p_{X^n}(\mathcal{S}_\gamma^c) + \frac{1}{\sqrt{n}}, \qquad (14)$$

where we have used the theorem with X^n being the $X_{\mathcal{V}}$ in the statement of the theorem. Further

$$\mathcal{S}_\gamma := \{x^n : h_{p_{X^n}}(x^n) - \log n > n(R + \tilde{R})\}. \qquad (15)$$

Note that for each $x^n \in \mathcal{T}_{\Phi_X^{(n)}}$ the relation $h_{p_{X^n}}(x^n) = \log|\mathcal{T}_{\Phi_X^{(n)}}|$ holds. Hence if we set

$$n(R + \tilde{R}) = nH_{\Phi_X^{(n)}}(X) - (L + 2)\log n, \qquad (16)$$

then the first term of (14) is zero and we have $\epsilon_{\mathsf{Apx}} \leq \dfrac{1}{\sqrt{n}}$.

Next we should find $\tilde{R}$ such that the error probability $\epsilon_{\mathsf{Dec}} \leq \epsilon - \frac{1}{\sqrt{n}}$. The decoder has access to Y^n and a single bin index F of X^n. Setting $\gamma = \dfrac{1}{2}\log n$, $|\mathcal{V}| = 1$, $Z = Y^n$ as well as F as a bin index of $X_{\mathcal{V}} = X^n$ in the statement of Theorem 2, we get that for any $t_{X^nY^n}$, we have

$$\epsilon_{\mathsf{Dec}} \leq p_{X^n}q_{Y^n|X^n}(\mathcal{S}(t)^c) + \frac{1}{\sqrt{n}}, \tag{17}$$

where $\mathcal{S}(t) := \{(x^n, y^n) : n\tilde{R} - h_t(x^n|y^n) > \dfrac{1}{2}\log n\}$. Observe that $Y_1, \cdots, Y_n$ are conditionally independent given any $X^n = x^n$ because the channel is memoryless. So if we can write $h_t(x^n|y^n)$ as a sum of independent rv's, we would be able to use Berry-Esseen CLT to find $\tilde{R}$. Using $t_{X^nY^n} = p_{X^n}q_{Y^n|X^n}$ does not give rise to such a factorization. To overcome this situation we use $t_{X^n,Y^n} = q_{X^n}q_{Y^n|X^n} = \prod_{i=1}^n q(x_i)q(y_i|x_i)$ for mismatch decoding. We then have $h_t(x^n|y^n) = \sum_{i=1}^n h_q(x_i|y_i)$. Given $X^n = x^n$, $\{h_q(x_i|Y_i)\}_{i=1}^n$ are functions of independent rv's, and hence mutually independent; thus we can now apply the Berry-Esseen CLT to bound the first term of (17). Using the Berry-Esseen CLT for each $x^n \in \mathcal{T}_{\Phi_X^{(n)}}$, we have

$$q_{Y^n|X^n=x^n}(\mathcal{S}(t)^c) = p_G(G \geq n\tilde{R} - \frac{1}{2}\log n) + O(\frac{1}{\sqrt{n}})$$

where G is a normal r.v. with

$$\begin{aligned}\mathbb{E}G &= \sum_{i=1}^n \mathbb{E}_{q_{Y_i|x_i}} h_q(x_i|Y_i) = \sum_x \#[x_i = x]\mathbb{E}_{q_{Y|x}} h_q(x|Y)\\ &= \sum_x n\Phi^{(n)}(x)\mathbb{E}_{q_{Y|x}} h_q(x|Y) = n\mathbb{E}_{\Phi_X^{(n)}}\mathbb{E}_{q_{Y|X}}[h_q(X|Y)|X]\end{aligned}$$

$$\begin{aligned}\mathsf{Var}G &= \sum_{i=1}^n \mathsf{Var}_{q_{Y_i|x_i}} h_q(x_i|Y_i) = n\mathbb{E}_{\Phi_X^{(n)}}\mathsf{Var}_{q_{Y|X}}[h_q(X|Y)|X]\\ &= n\mathbb{E}_{\Phi_X^{(n)}}\mathsf{Var}_{q_{Y|X}}[\imath_q(X;Y)|X].\end{aligned}$$

The sketch of the rest of the proof is as follows (see [10] for details): analyzing the bound (17), we get that

$$\begin{aligned}n\tilde{R} &= n\mathbb{E}_{\Phi_X^{(n)}}\mathbb{E}_{q_{Y|X}}[h_q(X|Y)|X]+\\ &\sqrt{n\mathbb{E}_{\Phi_X^{(n)}}\mathsf{Var}_{q_{Y|X}}[\imath_q(X;Y)|X]}Q^{-1}(\epsilon) + O(\log n),\end{aligned} \tag{18}$$

is sufficient to achieve $\epsilon_{\mathsf{Dec}} \leq \epsilon - \frac{1}{\sqrt{n}}$. Finally combining (16), (18) with $\|\Phi_X^{(n)} - q_X\|_\infty \leq \frac{1}{n}$ imply (12).

B. Broadcast channel

Consider the problem of transmission of two private messages over a broadcast channel $q_{Y_1Y_2|X}$. Let $\mathcal{R}^*(n, \epsilon)$ be the set of all rate pairs (R_1, R_2) of all (n, ϵ)-codes where ϵ is the probability of erroneous decoding at either of the decoders. 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 [10].

Theorem 3: Given any pmf $q_{U_1U_2X}$, let $\mathcal{R}_{\mathsf{in}}(q_{U_1U_2X}, n, \epsilon)$ be the set of all pairs (R_1, R_2) for which there exists reals $\tilde{R}_1, \tilde{R}_2 \geq 0$ such that

$$R_j + \tilde{R}_j \leq H_q(U_j) - O(\frac{\log n}{n}),\ j = 1, 2,$$

$$R_1 + R_2 + \tilde{R}_1 + \tilde{R}_2 \leq H_q(U_1U_2) - O(\frac{\log n}{n}) \tag{19}$$

$$\begin{bmatrix}\tilde{R}_1\\ \tilde{R}_2\end{bmatrix} \in \begin{bmatrix}H_q(U_1|Y_1)\\ H_q(U_2|Y_2)\end{bmatrix} + \frac{1}{\sqrt{n}}\mathcal{Q}^{-1}(\mathbb{V}_{\mathsf{BC},q}, \epsilon) + O(\frac{\log n}{n}), \tag{20}$$

where the entropies are computed according to the pmf $q_{U_1U_2XY_1Y_2} = q_{U_1U_2X}q_{Y_1Y_2|X}$ and

$$\mathbb{V}_{\mathsf{BC},q} = \mathbb{E}_{q_{U_1U_2}}\mathsf{Cov}_{q_{Y_1Y_2|U_1U_2}}\left[(\imath_q(U_1;Y_1), \imath_q(U_2;Y_2))^{\mathsf{T}}|U_1U_2\right].$$

Then $\cup_{q_{U_1U_2X}}\mathcal{R}_{\mathsf{in}}(q_{U_1U_2X}, n, \epsilon) \subseteq \mathcal{R}^*(n, \epsilon)$.

Sketch of the proof: The proof follows in similar steps as in the proof of channel coding.

Part 1: Source coding side of the problem and random binning: Let (U_1, U_2, X, Y) be distributed according to $q(u_1, u_2, x, y_1, y_2) = q(u_1, u_2, x)q(y_1, y_2|x)$. Consider the following random binning

- For $j = 1, 2$, to each u_j assign independently two random bins $m_j \in [1 : \mathsf{M}_j]$ and $f_j \in [1 : \mathsf{F}_j]$.

Suppose that the decoder at the receiver $j = 1, 2$ uses a mismatched decoder $T_j(\hat{u}_j|y_j, f_j)$ to generate $\hat{u}_j$ and thereby $\hat{m}_j$. The induced random pmf is

$$\begin{aligned}P_s(u_{1:2}, m_{1:2}, f_{1:2}, y_{1:2}, \hat{u}_{1:2}) &= q(u_{1:2}, y_{1:2})P_s(m_1, f_1|u_1)\\ &P_s(m_2, f_2|u_2)T_1(\hat{u}_1|f_1, y_1)T_2(\hat{u}_2|f_2, y_2).\end{aligned} \tag{21}$$

We find an upper bound on $\mathbb{E}\mathsf{P}(\hat{u}_1 \neq u_1 \mathsf{or}\ \hat{u}_2 \neq u_2)$ which in turn bounds the probability of error. Using the first bound of Theorem 2 and the union bound, we have (see [10] for proof):

Lemma 1: For any mismatched decoders $T_j, j = 1, 2$,

$$\mathbb{E}\mathsf{P}[\mathcal{E}] \leq p_{U_{1:2}Y_{1:2}}(\mathcal{S}_\gamma(t_1, t_2)^c) + 4 \times 2^{-\gamma}, \tag{22}$$

where γ is an arbitrary positive number and

$$\mathcal{S}_\gamma(t_1, t_2) := \{(u_{1:2}, y_{1:2}) : \log \mathsf{F}_j - h_{t_j}(u_j|y_j) > \gamma, j = 1, 2\}.$$

Part 2: Designing encoder-decoder assisted with a shared randomness: Assume that there is a shared randomness (F_1, F_2) available at the both encoders and the decoder, which is independent of the message and uniformly distributed over $[1 : \mathsf{F}_1] \times [1 : \mathsf{F}_2]$. The encoder uses the conditional pmf $P_s(u_{1:2}, x|m_{1:2}, f_{1:2})$ of the source coding problem. The decoder j uses the mismatched decoder $T_j(\hat{u}_j|y_j, f_j)$ to find $\hat{u}_j$ and as a result an estimate $\hat{m}_j$ of the message. Then the induced random pmf is $P_c(u_{1:2}, x, y_{1:2}, m_{1:2}, f_{1:2}, \hat{u}_{1:2}) = p^U(m_{1:2}, f_{1:2})P_s(u_{1:2}, x, y_{1:2}, \hat{u}_{1:2}|m_{1:2}, f_{1:2})$. We have

$$\|P_s - P_c\|_1 = \left\|P_s(m_{1:2}, f_{1:2}) - p^U(m_{1:2}, f_{1:2})\right\|_1.$$

The probability of error is no more than $\|P_s - P_c\|_1$ and thus no more than the right hand side of the above equation. Given M and F, Theorem 1 gives an upper bound ϵ_{Apx} on the expectation of the right hand side. Observe that the expected error probability $\mathbb{E}_{\mathcal{B}}\mathsf{P}[\mathcal{E}]$ of the channel coding is bounded from above by $\epsilon_{\mathsf{Dec}} + \epsilon_{\mathsf{Apx}}$.

Finally we can eliminate the shared randomness $F_{1:2}$ as in the proof of the channel coding.

Step 2: Non-asymptotic analysis of one-shot OSRB: We would apply the one shot OSRB bound to n repetitions of the BC $q_{Y_{1:2}|X}$.

Let $\mathsf{M}_j = 2^{nR_j}$ and $\mathsf{F}_j = 2^{n\tilde{R}_j}$. Following the proof of channel coding, for a given q_X and n, we find an n-type $\Phi^{(n)}_{U_{1:2}}$ such that $\|\Phi^{(n)}_{U_{1:2}} - q_{U_{1:2}}\|_\infty \le \frac{1}{n}$. To prove (12), assume that $p_{U^n_{1:2}}$ is a uniform distribution over the set $\mathcal{T}_{\Phi^{(n)}_{U_{1:2}}}$ of sequences with the type $\Phi^{(n)}_{U_{1:2}}$. Observe that $p_{U^n_j}$ has a uniform distribution over the set $\mathcal{T}_{\Phi^{(n)}_{U_j}}$ of sequences with the type $\Phi^{(n)}_{U_j}$. As a result, $h_{p_{U^n_{1:2}}}(u^n_j) = \log|\mathcal{T}_{\Phi^{(n)}_{U_j}}|$. As in the proof of channel coding, we can utilize Theorem 1 to show that if the inequalities in (19) are satisfied, then $\epsilon_{\mathsf{Apx}} \le O(\dfrac{1}{\sqrt{n}})$.

Next we should find $\tilde{R}_1, \tilde{R}_2$ such that $\epsilon_{\mathsf{Dec}} \le \epsilon - O(\frac{1}{\sqrt{n}})$. We can utilize Lemma 1 to show that (20) is sufficient for $\epsilon_{\mathsf{Dec}} \le \epsilon - O(\frac{1}{\sqrt{n}})$. The rest of the proof is similar to that of channel coding but uses a generalized version of Berry-Esseen CLT for the independent and multidimensional r.v.'s [9]. ■

C. Wiretap channel with strong secrecy

Consider a wiretap channel with probability transition $q_{YZ|X}$, in which the receiver and the wiretapper have access to channel outputs Y and Z, respectively. For a given (n, R) code we use total variation distance $\left\|p_{MZ^n} - p^U_M p_{Z^n}\right\|_1$ to measure the security of the code, where p_{MZ^n} is the induced pmf by the code. A rate R is said to be $(\epsilon_r, \epsilon_{\mathsf{sec}})$-achievable if there exists an (n, R) code such that $\mathsf{P}[\mathcal{E}] \le \epsilon_r$ and $\left\|p_{MZ^n} - p^U_M p_{Z^n}\right\|_1 \le \epsilon_{\mathsf{sec}}$.

Theorem 4: Given $q_{Y,Z|X}$, for any input distribution $q_{U,X}$ and any $\theta \in [0, 1]$, the following rate is $(n, \epsilon_r, \epsilon_{\mathsf{sec}})$-achievable:

$$R(n,\epsilon_r, \epsilon_{\mathsf{sec}}) = R_1 - \frac{1}{\sqrt{n}}R_2 - O(\frac{\log n}{n}) \tag{23}$$

where $R_1 = I_q(U;Y) - I_q(U;Z)$ and

$$R_2 = \sqrt{\mathsf{V}_Y}Q^{-1}(\theta\epsilon_r) + \sqrt{\mathsf{V}_Z}Q^{-1}(\bar{\theta}\epsilon_{\mathsf{sec}}),$$

in which $\bar{\theta} = 1 - \theta$, $\mathsf{V}_Y = \mathbb{E}_{q_{UX}}\mathsf{Var}_{q_{Y|UX}}[I_q(U;Y)|U]$ and V_Z is defined similarly.

Remark 4: Tan in [8] obtains the second order coding rate $R_2 = \sqrt{\mathsf{V}_Y}Q^{-1}(\frac{\epsilon_r}{2}) + \sqrt{\mathsf{V}_Z}Q^{-1}(\frac{\epsilon^2_{\mathsf{sec}}}{400})$ which is bigger than the one obtained in Theorem 4 for $\theta = \frac{1}{2}$.

Sketch of proof: For simplicity, we prove the theorem for the special case $U = X$. We will find $R(n, \epsilon_r, \epsilon_{\mathsf{sec}})$ such that $\mathbb{E}\mathsf{P}[\mathcal{E}] \le \theta\epsilon_r$ and $\mathbb{E}\left\|p_{MZ^n} - p^U_M p_{Z^n}\right\|_1 < \bar{\theta}\epsilon_{\mathsf{sec}}$. Then by Markov inequality, we can find a code with the desired conditions.

One-shot OSRB: We use the same code construction of subsection IV-A. Here we need to compute the security index of the code. To do this, we bound the security constraint, i.e. $\epsilon_{s,\mathsf{sec}} = \mathbb{E}\left\|P_s(m, f, z) - p^U(m, f)q(z)\right\|_1$, using Theorem 1 in the source coding part of the problem. Then using triangle inequality, the security constraint of channel coding asserted with shared randomness $\epsilon_{c,\mathsf{sec}} = \mathbb{E}\left\|P_c(m, f, z) - p^U(m, f)q(z)\right\|_1$ is bounded above by $\epsilon_{s,\mathsf{sec}} + \epsilon_{\mathsf{Apx}}$. To eliminate F, we can show that there exists an instance f such that $\mathbb{E}_{\mathcal{B}|F=f}\left\|P_c(m, z|f) - p^U(m)P_c(z)\right\|_1 \le \epsilon_{s,\mathsf{sec}} + 3\epsilon_{\mathsf{Apx}}$.

Non-asymptotic analysis of one-shot OSRB: Again we follow the analysis of pt-to-pt channel. It was observed that if (16) is satisfied, then $\epsilon_{\mathsf{Apx}} \le 1/\sqrt{n}$. Similar error analysis shows that $\mathbb{E}\mathsf{P}[\mathcal{E}] \le \theta\epsilon_r$ provided that

$$n\tilde{R} = nH_q(X|Y) + \sqrt{n\mathsf{V}_Y}Q^{-1}(\theta\epsilon_r) + O(\log n). \tag{24}$$

Next we find a constraint on R and $\tilde{R}$ such that security index $\epsilon_{s,\mathsf{sec}} \le \bar{\theta}\epsilon_{\mathsf{sec}} - 3/\sqrt{n}$, which shows that $\mathbb{E}\left\|p_{MZ^n} - p^U_M p_{Z^n}\right\|_1 < \bar{\theta}\epsilon_{\mathsf{sec}}$. Substituting $\gamma = \log n$ in Theorem 1 gives: $\epsilon_{s,\mathsf{sec}} \le p_{X^n}(\mathcal{S}^c_\gamma(p\|t)) + 1/\sqrt{n}$, where

$$\mathcal{S}_\gamma(p\|t) := \{(x^n, z^n) : h_{p_{X^n}}(x^n, z^n) - h_t(z^n) - \log n > n(R + \tilde{R})\}.$$

Again as in the proof of error probability for channel coding, to apply Berry-Esseen CLT we need to write $h_{p_{X^n}}(x^n, z^n) - h_t(z^n)$ as a sum of independent r.v.'s. Let $t(z^n) = \prod_{i=1}^n q_Z(z_i)$, which makes $h_t(z^n)$ as sum of independent r.v.'s. Next observe that for any $x^n \in \mathcal{T}_{\Phi_{X^n}}$, $p_{X^n}q_{Z^n|X^n}(x^n, z^n) = q_{X^n,Z^n}(x^n, z^n)2^{-O(\log n)}$. Using this fact, we have $h_{p_{X^n}q_{Z^n|X^n}}(x^n, z^n) = h_q(x^n, z^n) + O(\log n)$; thus

$$\mathcal{S}_\gamma(p\|t) := \{(x^n, z^n) : \sum_{i=1}^n h_q(x_i|z_i) - O(\log n) > n(R + \tilde{R})\}.$$

Applying Berry-Esseen CLT in the same way as in channel coding proof implies

$$n(R + \tilde{R}) = nH_q(X|Z) - \sqrt{n\mathsf{V}_Z}Q^{-1}(\bar{\theta}\epsilon_{\mathsf{sec}}) + O(\log n). \tag{25}$$

Combining (24) and (25) yields (23). ■

References

[1] M. H. Yassaee, M. R. Aref and A. Gohari, "Achievability proof via output statistics of random binning," arXiv:1203.0730, IEEE Symp. On IT (ISIT) 2012, pp. 1049-1054.

[2] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informations theorie", in Trans. Third. Prague Conf. Inf. Theory, 1962, pp. 689723.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding in the finite blocklength regime", *IEEE Trans. Inf. Theory*, 56(5), 2307 59, 2010.

[4] S. Verdú, "Non-Asymptotic Achievability Bounds in Multiuser Information Theory", Allerton Conference, Oct. 2012.

[5] V. Kostina, S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime", arXiv:1209.1317, Sep. 2012.

[6] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding", in Allerton Conference, 2011, arXiv:1109.6310.

[7] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems", arXiv:1201.3901, Feb 2012.

[8] V. Y. F. Tan, "Achievable Second-Order Coding Rates for the Wiretap Channel", IEEE Int. Conf. on Comm. Systems (ICCS) 2012, Singapore.

[9] V. Bentkus, "A Lyapunov-type bound in $\mathbb{R}^d$," Theory of Probability & Its Applications, 49(2), 311-323, 2005.

[10] M. H. Yassaee, M. R. Aref and A. Gohari, "Non-asymptotic output statistics of random binning and its applications", arXiv:1303.0695.

[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.