Synthesizing Correlated Randomness using Algebraic Structured Codes

Touheed Anwar Atif
University of Michigan, USA
Email: touheed@umich.edu

Arun Padakandla
University of Tennessee, USA
Email: arunpr@utk.edu

S. Sandeep Pradhan
University of Michigan, USA
Email: pradhanv@umich.edu

***Abstract*—In this problem, Alice and Bob, are provided X_1^n and X_2^n that are IID $p_{X_1X_2}$. Alice and Bob can communicate to Charles over (noiseless) links of rate R_1 and R_2, respectively. Their goal is to enable Charles generate samples Y^n such that the triple (X_1^n, X_2^n, Y^n) has a PMF that is close, in total variation, to $\prod p_{X_1X_2Y}$. In addition, the three parties may posses shared common randomness at rate C. We address the problem of characterizing the set of rate triples (R_1, R_2, C) for which the above goal can be accomplished. We build on our recent findings and propose a new coding scheme based on coset codes. We analyze its information-theoretic performance and derive a new inner bound. We identify examples for which the derived inner bound is analytically proven to contain rate triples that are not achievable via any known unstructured code based coding techniques. Our findings build on a variant of soft-covering which generalizes its applicability to the algebraic structured code ensembles. This adds to the advancement of the use structured codes in network information theory.**

I. Introduction

We consider the scenario which was originally studied by authors in [1], as depicted in Fig 1. Three distributed parties, say Alice, Bob and Charles, have to generate samples that are independent and identically distributed (IID) with a target probability mass function (PMF) $p_{X_1X_2Y}$. Alice and Bob are provided with samples that are IID according to $p_{X_1X_2}$ - the marginal of the target PMF $p_{X_1X_2Y}$. They have access to unlimited private randomness and share noiseless communication links of rates R_1, R_2 with Charles. In addition, the three parties share common randomness at rate C. The authors in [1] provided a set of sufficient conditions, i.e., an achievable rate region for such a scenario. However, can this rate-region be improved? This article answers the above question in the affirmative.

It is well established that traditional coding techniques using unstructured codes do not achieve optimality for the several multi-terminal scenarios. For instance, the work by Körner-Marton [2] demonstrated this sub-optimality for a classical distributed lossless compression problem with symmetric binary sources using random linear codes. We harness analogous gains for the problem of generating correlated randomness at distributed parties. Specifically, we propose a coding scheme based on coset codes, analyze its information-theoretic performance and thereby derive a new inner bound (see Thm. 1). We identify an example for which the derived inner bound is analytically proven to contain rate triples that are not achievable in the earlier known results [1]. While the derived inner bound does not subsume the one characterized in [1], one can adopt the technique in [3, Sec. VII] - also demonstrated in a related context [4] - to derive an inner bound that subsumes the inner bounds derived in [1] and Thm. 1.

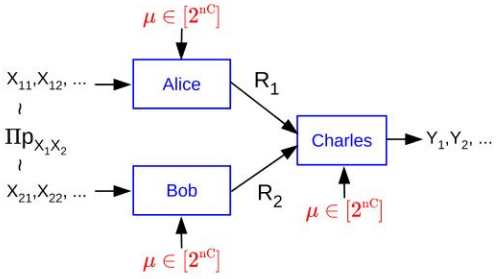

Fig. 1. Source Coding for Synthesizing Correlated Randomness

The problem of generating correlated randomness can be traced back to Wyner [5], whose work discovered the important technical tool, called the *soft covering*. This tool has found its application in diverse fields including cryptography and quantum information theory. The work in [1] further refined this tool by introducing a joint-typicality based application. As we illustrate in the sequel, this work adds another dimension to our current understanding of soft covering, what we term as the *change of measure soft covering*.

A renewed interest in soft covering led Cuff [6], [7] to consider a point-to-point (PTP) version of the scenario depicted in Fig. 1, wherein Bob (or X_2) is absent. A side-information based scenario was subsequently studied in [8] and a converse provided in [1]. In [1] we studied the above scenario using unstructured coding techniques. A similar sequence of problems were also studied in the quantum setting [9]–[11].

While all of the above works leverage the unstructured IID random codes, it has been proven that algebraic structured codes provide gains in network communication involving distributed encoders [4], [12]–[17]. Motivated by this, we consider the distributed correlation synthesis problem depicted in Fig. 1 and present a new achievable rate-region using structured coding techniques. We highlight two main challenges in this endeavour. The first challenge is to be able to achieve rates corresponding to non-uniform distributions. In particular, codewords within a random linear code has uniform empirical distributions. This requires us to enlarge our codes to be able to identify codeword with the desired single-letter distribution. We address this challenge by using a random shifts of cosets of a linear code as our code, henceforth referred to as Unionized Coset Codes (UCCs) [16]. The second challenge concerns the statistical dependence among codewords of a coset code. In contrast to IID codes, the codewords of a UCC are only

This work was supported by NSF grant CCF-2007878.

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517918

pairwise independent [18]. This prevents us from using the Chernoff concentration bound. We therefore develop novel techniques for our information theoretic study.

II. Preliminaries and Problem Statement

We supplement standard information theory notation with the following. For a PMF p_X, we let $p_X^n = \prod_{i=1}^n p_X$. For an integer $n \geqslant 1$, $[n] \triangleq \{1, \cdots, n\}$. The total variation between PMFs p_X and q_X defined over $\mathcal{X}$ is denoted $\|p_X - q_X\|_1 = \frac{1}{2}\sum_{x\in\mathcal{X}} |p_X(x) - q_X(x)|$. $\mathbb{F}_p$ is used to denote a finite field of size p with addition $\oplus$.

Building on this, we address the network scenario (Fig. 1) for which we state the problem below. In the following, we let $\underline{X} = (X_1, X_2), \underline{x}^n = (x_1^n, x_2^n)$.

Definition 1. Given a PMF $p_{X_1X_2Y}$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$, a rate triple (R_1, R_2, C) is achievable, if $\forall \epsilon > 0$ and sufficiently large n, there exists 2^{nC} randomized encoder pairs $E_j^{(\mu)} : \mathcal{X}_j^n \to [\Theta_j] : j \in [2], \mu \in [2^{nC}]$, and a corresponding collection of 2^{nC} randomized decoders $D^{(\mu)} : [\Theta_1] \times [\Theta_2] \to \mathcal{Y}^n$ for $\mu \in [2^{nC}]$ such that $\left| p^n_{\underline{X}Y} - p_{\underline{X}^n Y^n} \right|_1 \leqslant \epsilon$, $\frac{1}{n}\log_2 \Theta_j \leqslant R_j + \epsilon : j \in [2]$, where

$$p_{\underline{X}^n Y^n}(\underline{x}^n, y^n) = \sum_{\mu\in[2^{nC}]} 2^{-nC} \sum_{\substack{(m_1,m_2)\in \\ [\Theta_1]\times[\Theta_2]}} p^n_{\underline{X}Y}(\underline{x}^n, y^n)$$
$$p^{(\mu)}_{M_1|X_1^n}(m_1|x_1^n) p^{(\mu)}_{M_2|X_2^n}(m_2|x_2^n) p^{(\mu)}_{Y^n|M_1,M_2}(y^n|m_1, m_2)$$

$p^{(\mu)}_{M_j|X_j^n} : j \in [2], p^{(\mu)}_{Y^n|M_1,M_2}$ are the PMFs induced by the two randomized encoders and decoder respectively, corresponding to common randomness message μ. We let $\mathcal{R}_d(p_{\underline{X}Y})$ denote the set of achievable rate triples.

Theorem 1 provides a new characterization of $\mathcal{R}_d(p_{\underline{X}Y})$ based on coset codes, for the above described problem statement. This characterization provides a new inner bound to the achievable rate-region. An essential aspect of our work is the identification of a PMF $p_{X_1X_2Y}$ for which the coding scheme described in [1], [19] is strictly sub-optimal.

III. Distributed Soft Covering using Algebraic Structured Random Codes

A. Change of Measure Soft Covering

Before presenting the main result of the paper, we develop the necessary tools and provide a lemma which is crucial for the upcoming results. This lemma extends the cloud mixing result of [7] with a mismatched codebook generation process. The lemma is as follows.

Lemma 1. *Consider a PMF p_{XY} on $\mathcal{X} \times \mathcal{Y}$, and let R be a finite non-negative integer. Additionally, assume that there exists some set $\bar{\mathcal{X}}$ containing the set $\mathcal{X}$, with $p_{XY}(x, y) = 0$ for all $x \in \bar{\mathcal{X}} \backslash \mathcal{X}$. Suppose q_X is any PMF on the set $\bar{\mathcal{X}}$ such that the PMF p_X is absolutely continuous with respect to the q_X. Let a random code $\mathbb{C} \triangleq \{X^n(m) : m \in [2^{nR}]\}$ be defined as a collection of codewords chosen pairwise independently from the set $\bar{\mathcal{X}}$ according to the PMF q_X^n. Then we have for $R \geqslant H_q(X) - H_p(Y|X) = I_p(X;Y) - H_p(X) + H_q(X)$,*

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{C}} \left[\sum_{y^n\in\mathcal{Y}^n} \left| p_Y^n(y^n) - \frac{1}{M}\sum_{m=1}^{2^{nR}} \frac{p_X^n(X^n(m))}{q_X^n(X^n(m))} p_{Y|X}^n(y^n|X^n(m)) \right| \right] = 0$$

Proof. The proof is provided in Appendix of [20]. □

B. Main Result

Our main result is the characterization of $\mathcal{R}_s(p_{\underline{X}Y})$ which is the inner bound to $\mathcal{R}_d(p_{\underline{X}Y})$. In the following, we let $\underline{X} = (X_1, X_2), \underline{W} = (W_1, W_2), \underline{x} = (x_1, x_2)$ and $\underline{w} = (w_1, w_2)$.

Theorem 1. *Given a PMF $p_{X_1X_2Y}$, let $\mathcal{P}(p_{X_1X_2Y})$ denote the collection of all PMFs $p_{QW_1W_2\underline{X}Y}$ defined on $\mathcal{Q}\times\mathcal{W}_1\times\mathcal{W}_2\times\underline{\mathcal{X}}\times\mathcal{Y}$ such that (i) $p_{\underline{X}Y}(\underline{x}, y) = \sum_{(q,\underline{w})\in\mathcal{Q}\times\underline{\mathcal{W}}} p_{Q\underline{W}\underline{X}Y}(q, \underline{w}, \underline{x}, y)$ for all $(\underline{x}, y) \in \underline{\mathcal{X}} \times \mathcal{Y}$, (ii) $W_1 - Q\underline{X} - QX_2 - W_2$ and $\underline{X} - Q\underline{W} - Y$ are Markov chains, (iii) $|\mathcal{W}_1| \leqslant |\mathcal{X}_1|$, $|\mathcal{W}_2| \leqslant |\mathcal{X}_2|$. Further, let $\beta(p_{Q\underline{W}\underline{X}Y})$ denote the set of rates and common randomness triple (R_1, R_2, C) that satisfy*

$$\begin{aligned}
R_1 &\geqslant I(X_1; W_1|W_2, Q) + I(W_1 \oplus W_2; W_2|Q)\\
R_2 &\geqslant I(X_2; W_2|W_1, Q) + I(W_1 \oplus W_2; W_1|Q)\\
R_1 + C &\geqslant I(\underline{X}; W_1|W_2, Q) + I(Y; W_1|\underline{X}, Q)\\
&\quad + I(W_1 \oplus W_2; W_2|Q)\\
R_2 + C &\geqslant I(\underline{X}; W_2|W_1, Q) + I(Y; W_2|\underline{X}, Q)\\
&\quad + I(W_1 \oplus W_2; W_1|Q)\\
R_1 + R_2 + C &\geqslant I(\underline{X}; W_1|W_2, Q) + I(\underline{X}; W_2|W_1, Q)\\
&\quad + I(W_1 \oplus W_2; W_1|Q) + I(W_1 \oplus W_2; W_2|Q)
\end{aligned} \tag{1}$$

where the above information theoretic terms are evaluated with respect to the PMF $p_{QW_1W_2\underline{X}Y}$. Let

$$\mathcal{R}_s(p_{\underline{X}Y}) \triangleq \textit{Closure}\left(\bigcup_{p_{Q\underline{W}\underline{X}Y}\in\mathcal{P}(p_{X_1X_2Y})} \beta(p_{Q\underline{W}\underline{X}Y}) \right) \tag{2}$$

We have

$$\mathcal{R}_s(p_{\underline{X}Y}) \subseteq \mathcal{R}_d(p_{\underline{X}Y}).$$

In other words, the rate triple $(R_1, R_2, C) \in \left(\bigcup_{p_{Q\underline{W}\underline{X}Y}\in\mathcal{P}(p_{X_1X_2Y})} \beta(p_{Q\underline{W}\underline{X}Y})\right)$ is achievable.

Note that the rate-region obtained in Theorem 2 of [19] contains the constraint $R_1 + R_2 + C \geqslant I(X_1X_2Y; W_1W_2|Q)$. Hence when $2H(W_1 \oplus W_2|Q) < H(W_1, W_2|Q)$, the above theorem gives a lower sum rate constraint. As a result, the rate-region above contains points that are not contained within the rate-region provided in [19]. To illustrate this fact further, consider the following example.

Example 1. Let X_1 and X_2 be a pair of binary symmetric correlated sources with $P(X_2 = 1|X_1 = 0) = p$, for some $p \in (0, 0.5)$. Let $Y = X_1 \oplus X_2 \oplus Q$, where $P(Q = 1) = q$,

for some $q \in (0, 0.5)$. Consider $q = p = 0.1$ for a numerical evaluation. Let us first consider the inner bound $\mathcal{R}_u(p_{\underline{X},Y})$ to the rate region $\mathcal{R}(p_{\underline{X},Y})$ given in [1], developed using unstructured code ensemble. Due to symmetry in the example, it turns out that the search over the auxiliary random variables for minimization reduces to a single-parameter minimization which can be computed through derivative techniques. The computation details are not provided for the sake of brevity. In particular, the minimum value of $R_1 + R_2 + C$ can be computed to be 1.3965. Next let us consider the new inner bound $\mathcal{R}_s(p_{\underline{X},Y})$ developed using structured code ensemble (Theorem 1). The minimum value of $R_1 + R_2 + C$ can be computed to be 0.9596.

The results can also be verified for the special case of $q = 0$ which we provide in the following. Using the arguments given in proof of Proposition 1 of [2], one can show that

$$\mathcal{R}_u(p_{\underline{X},Y}) = \{(R_1, R_2, C) : R_1 \geqslant h_b(p), R_2 \geqslant h_b(p), \\ R_1 + R_2 \geqslant 1 + h_b(p), C \geqslant 0\}.$$

Next let us consider the new inner bound $\mathcal{R}_s(p_{\underline{X},Y})$ developed using structured code ensemble (Theorem 1). By choosing $W_1 = X_1$ and $W_2 = X_2$, we see that the following triple of rates is achievable:

$$\{(R_1, R_2, C) : R_1 \geqslant h_b(p), R_2 \geqslant h_b(p), C \geqslant 0\}.$$

In fact, one can show that this is optimal using the side information argument. If X_2 is sent losslessly, then from the converse argument in the side information case, we see that $R_1 \geqslant H(X_2|X_1) = h_b(p)$.

IV. Proof of Distributed Soft Covering using Algebraic Structured Random Codes

The coding strategy used here is based on Unionized Coset Codes, defined in Definition (2). The structure in these codes provides a method to exploit the structure present in the stochastic processing applied by decoder, i.e., $P_{Y|W_1+W_2}$. Using this technique, we aim to strictly reduce the rate constraints compared to the ones obtained in Theorem 1 of [1].

Let $\mu \in [2^{nC}]$ denote the common randomness shared amidst all terminals. The first encoder uses a part of the entire common randomness available to it, say C_1 bits out of the C bits, which is denoted by $\mu_1 \in [2^{nC_1}]$. Similarly, let $\mu_2 \in [2^{nC_2}]$ denote the common randomness used by the second encoder. Our goal is to prove the existence of PMFs $p^{(\mu_1)}_{M_1|X_1^n}(m_1|x_1^n) : x_1^n \in \mathcal{X}_1^n, m_1 \in [\Theta_1], \mu_1 \in [2^{nC_1}]$, $p^{\mu_2}_{M_2|X_2^n}(m_2|x_2^n) : x_2^n \in \mathcal{X}_2^n, m_2 \in [\Theta_2], \mu_2 \in [2^{nC_2}]$, $p_{Y^n|M_1,M_2}(y^n|m_1, m_2) : y^n \in \mathcal{Y}^n, (m_1, m_2) \in [\Theta_1] \times [\Theta_2]$ such that

$$\mathscr{Q} \triangleq \frac{1}{2} \sum_{\underline{x}^n, y^n} \Bigg| p^n_{\underline{X}Y}(\underline{x}^n, y^n) - \sum_{\mu \in [2^{nC}]} \sum_{\substack{m_1 \in [\Theta_1], \\ m_2 \in [\Theta_2]}} \frac{p^n_{\underline{X}}(\underline{x}^n)}{2^{nC}} \\ p^{(\mu_1)}_{M_1|X_1^n}(m_1|x_1^n) p^{(\mu_2)}_{M_2|X_2^n}(m_2|x_2^n) p^{(\mu)}_{Y^n|\underline{M}}(y^n|\underline{m}) \Bigg| \leqslant \varepsilon,$$

$$\frac{\log \Theta_j}{n} \leqslant R_j + \epsilon : j \in [2], \tag{3}$$

for sufficiently large n. Fix a block length $n > 0$, a positive integer N and a finite field $\mathbb{F}_p$. Further, let W_1 and W_2 be random variables defined on the alphabets $\mathcal{W}_1$ and $\mathcal{W}_2$, respectively, where $\mathcal{W}_1 = \mathcal{W}_2 = \mathbb{F}_p$, and let $Z \triangleq W_1 \oplus W_2$. In building the code, we use the Unionized Coset Codes (UCCs) [16] defined as below. These codes involve two layers of codes (i) a coarse code and (ii) a fine code. The coarse code is a coset of the linear code and the fine code is the union of several cosets of the linear code.

For a fixed $k \times n$ matrix $G \in \mathbb{F}_p^{k \times n}$ with $k \leqslant n$, and a $1 \times n$ vector $B \in \mathbb{F}_p^n$, define the coset code as

$$\mathbb{C}(G, B) \triangleq \{x^n : x^n = a^k G + B, \text{ for some } a^k \in \mathbb{F}_p^k\}.$$

In other words, $\mathbb{C}(G, B)$ is a shift of the row space of the matrix G. The row space of G is a linear code. If the rank of G is k, then there are p^k codewords in the coset code.

Definition 2. An (n, k, l, p) UCC is a pair (G, h) consisting of a $k \times n$ matrix $G \in \mathbb{F}_p^{k \times n}$, and a mapping $h : \mathbb{F}_p^l \to \mathbb{F}_p^n$. In the context of UCC, define the composite code as $\mathbb{C} = \bigcup_{i \in \mathbb{F}_p^l} \mathbb{C}(G, h(i))$.

For every $\mu \triangleq (\mu_1, \mu_2)$, consider two UCCs $(G, h_1^{(\mu_1)})$ and $(G, h_2^{(\mu_2)})$, each with parameters (n, k, l_1, p) and (n, k, l_2, p), respectively. Note that, for every $\mu \in [N]$, the generator matrix G remains the same.

For each (μ_1, μ_2), the generator matrix G along with the function $h_1^{\mu_1}$ and $h_2^{\mu_2}$ generates p^{k+l_1} and p^{k+l_2} codewords, respectively. Each of these codewords are characterized by a triple (a_i, m_i, μ_i), where $a_i \in \mathbb{F}_p^k$ and $m_i \in \mathbb{F}_p^{l_i}$ corresponds to the coarse code and the fine code indices, respectively, for $i \in [2]$. Let $\mathtt{w}_1(a_1, m_1, \mu_1)$ and $\mathtt{w}_2(a_2, m_2, \mu_2)$ denote the codewords associated with Alice and Bob, generated using the above procedure, respectively, where $\mathtt{w}_1(a_1, m_1, \mu_1) \triangleq a_1 G + h_1^{(\mu_1)}(i)$, and $\mathtt{w}_2(a_2, m_2, \mu_2) \triangleq a_2 G + h_2^{(\mu_2)}(j)$.

Consider the collections $c_1 = (c_1^{(\mu_1)} : 1 \leqslant \mu_1 \leqslant 2^{nC_1})$ where $c_1^{(\mu_1)} = (\mathtt{w}_1(l_1, \mu_1) : 1 \leqslant l_1 \leqslant 2^{n\tilde{R}_1})$ and $c_2 = (c_1^{(\mu_1)} : 1 \leqslant \mu_1 \leqslant 2^{nC_1})$ where $c_2^{(\mu_2)} = (\mathtt{w}_2(l_2, \mu_2) : 1 \leqslant l_2 \leqslant 2^{n\tilde{R}_2})$. For this collection, we let

$$E^{(\mu_1)}_{L_1|X_1^n}(a_1, m_1|x_1^n) \triangleq \sum_{\substack{w_1^n \in \\ T_\delta(W_1|x_1^n)}} p^n \frac{p^n_{W_1|X_1}(w_1^n|x_1^n)}{2^{nS_1}(1+\eta)} \mathbb{1}_{\{\mathtt{w}_1(a_1, m_1, \mu_1) = w_1^n\}},$$

$$E^{(\mu_2)}_{L_2|X_2^n}(a_2, m_2|x_2^n) \triangleq \sum_{\substack{w_2^n \in \\ T_\delta(W_2|x_2^n)}} p^n \frac{p^n_{W_2|X_2}(w_2^n|x_2^n)}{2^{nS_2}(1+\eta)} \mathbb{1}_{\{\mathtt{w}_2(a_2, m_2, \mu_2) = w_2^n\}}.$$

The definition of $E^{(\mu_1)}_{L_1|X_1^n}$ and $E^{(\mu_2)}_{L_2|X_2^n}$ can be thought of as encoding rules that do not exploit the additional rebate obtained by using binning techniques, specifically in a distributed setup.

A. Binning of Random Encoders

We next proceed to binning the above constructed collection of random encoders. Since, UCC is already a union of several

cosets, we associate a bin to each coset, and place all the codewords of a coset in the same bin. For each $i \in \mathbb{F}_p^{l_1}$ and $j \in \mathbb{F}_p^{l_2}$, let $\mathcal{B}_1^{(\mu_1)}(i) \triangleq \mathbb{C}(G, h_1^{(\mu_1)}(i))$ and $\mathcal{B}_2^{(\mu_2)}(j) \triangleq \mathbb{C}(G, h_2^{(\mu_2)}(j))$ denote the i^{th} and the j^{th} bins, respectively. Formally, we define the following PMFs. $p^{(\mu_i)}_{M_i|X_i^n}(m_i|x_i^n)$

$$= \begin{cases} \mathbb{1}_{\{m_i=0\}} & \text{if } s_i^{(\mu_i)}(x_i^n) > 1, \\ 1 - s_i^{(\mu_i)}(x_i^n) & \text{if } m_i = 0 \text{ and } s_i^{(\mu_i)}(x_i^n) \in [0,1], \\ \sum_{a_i \in \mathbb{F}_p^k} E^{(\mu_i)}_{L_i|X_i^n}(a_i, m_i|x_i^n) & \text{if } m_i \neq 0 \text{ and } s_i^{(\mu_i)}(x_i^n) \in [0,1], \end{cases}$$

for all $x_1^n \in T_\delta(X_1)$, $s_1^{(\mu_1)}(x_1^n)$ defined as $s_1^{(\mu_1)}(x_1^n) \triangleq \sum_{a_1 \in \mathbb{F}_p^k} \sum_{m_1 \in \mathbb{F}_p^{l_1}} E^{(\mu_1)}_{a_1, m_1|X_1^n}(l_1|x_1^n)$ and $i \in [2]$.

With this definition note that, $\sum_{m_1=0}^{2^{nR_1}} p^{(\mu_1)}_{M_1|X_1^n}(m_1|x_1^n) = 1$ for all $\mu_1 \in [2^{nC_1}]$ and $x_1^n \in \mathcal{X}_1^n$ and similarly, $\sum_{m_2=0}^{2^{nR_2}} p^{(\mu_2)}_{M_2|X_2^n}(m_2|x_2^n) = 1$ for all $\mu_2 \in [2^{nC_2}]$ and $x_2^n \in \mathcal{X}_2^n$.

Also, note that the effect of introducing binning (by defining the above PMFs) is in reducing the communication rates from (S_1, S_2) to (R_1, R_2), where $R_i = \frac{l_i}{n} \log p, i \in \{1,2\}$. Now, we move on to describing the decoder.

B. Decoder mapping

We create a decoder that takes as an input a pair of bin numbers and produces a sequence $W^n \in \mathbb{F}_p^n$. More precisely, we define a mapping $f^{(\mu)}$ for $\mu \triangleq (\mu_1, \mu_2)$, acting on the messages (m_1, m_2) as follows. On observing μ and the classical indices $(m_1, m_2) \in \mathbb{F}_p^{l_1} \times \mathbb{F}_p^{l_2}$ communicated by the encoder, the decoder constructs $D_{i,j}^{(\mu)} \triangleq \{\tilde{a} \in \mathbb{F}_p^k : \tilde{a}G + h_1^{(\mu_1)}(i) + h_2^{(\mu_2)}(j) \in \mathcal{T}_{\hat{\delta}}^{(n)}(Z)\}$, and $f^{(\mu)}(m_1, m_2)$

$$\triangleq \begin{cases} \tilde{a}G + h_1^{(\mu_1)}(i) + h_2^{(\mu_2)}(j) & \text{if } D_{i,j}^{(\mu_1,\mu_2)} \equiv \{\tilde{a}\} \\ w_0^n & \text{otherwise}, \end{cases} \tag{4}$$

where $\hat{\delta} = p\delta$ and w_0^n is an additional sequence added to $\mathbb{F}_p^n$. Further, $f^{(\mu)}(m_1, m_2) = w_0^n$ for $i = 0$ or $j = 0$. The decoder then performs a stochastic processing of the output and chooses y^n according to PMF $p^n_{Y|Z}(y^n|f^{(\mu)}(m_1, m_2))$. This implies the PMF $p^{(\mu_1)}_{Y^n|M_1M_2}(\cdot|\cdot)$ is given by

$$p^{(\mu)}_{Y^n|M_1M_2}(\cdot|m_1, m_2) = p^n_{Y|Z}(y^n|f^{(\mu)}(m_1, m_2)). \tag{5}$$

We now begin our analysis of the total variation term given in (3).

C. Analysis of Total Variation

Our goal is to prove the existence of a collections c_1, c_2 for which (3) holds. We do this via random coding. Specifically, we prove that $\mathbb{E}[K] \leqslant \epsilon$, where the expectation is over the ensemble of codebooks. The PMF induced on the ensemble of codebooks is as specified below. The codewords of the random codebook $C_i^{(\mu_i)} = (\mathtt{W}_i(a_i, m_i, \mu_i) : a_i \in \mathbb{F}_p^k, m_i \in \mathbb{F}_p^{l_i})$ for each $\mu_i \in [2^{nC_i}]$ are only pairwise independent [16] and distributed with PMF $\mathbb{P}(\mathtt{W}_i(a_i, m_i, \mu_i) = w_i^n) = \frac{1}{p^n}$ for each $i \in [2]$.

Step 1: Error caused by not covering

We begin by splitting K into two terms using the triangle inequality as $K \leqslant S + \tilde{S}$, where

$$S \triangleq \sum_{\underline{x}^n, y^n} \Bigg| p^n_{\underline{X}Y}(\underline{x}^n, y^n) - \sum_{\mu_1,\mu_2} \sum_{\substack{m_1>0, \\ m_2>0}} \frac{p^n_{\underline{X}}(\underline{x}^n) p^{(\mu_1)}_{M_1|X_1^n}(m_1|x_1^n)}{2^{n(C_1+C_2)}} p^{(\mu_2)}_{M_2|X_2^n}(m_2|x_2^n) p^{(\mu)}_{Y^n|\underline{M}}(y^n|\underline{m}) \Bigg|,$$

$$\tilde{S} \triangleq \sum_{\underline{x}^n, y^n} \Bigg| \sum_{\mu_1,\mu_2} \sum_{m_1=0 \bigcup m_2=0} \frac{p^n_{\underline{X}}(\underline{x}^n)}{2^{n(C_1+C_2)}} p^{(\mu_1)}_{M_1|X_1^n}(m_1|x_1^n) p^{(\mu_2)}_{M_2|X_2^n}(m_2|x_2^n) p^{(\mu)}_{Y^n|\underline{M}}(y^n|\underline{m}) \Bigg|.$$

Note that $\widetilde{S}$ captures the error induced by not covering $p^n_{\underline{X}Y}$. For the term corresponding to $\widetilde{S}$, we prove the following result.

Proposition 1. *There exist functions $\epsilon_{\widetilde{S}}(\delta)$, and $\delta_{\widetilde{S}}(\delta)$, such that for all sufficiently small δ and sufficiently large n, we have $\mathbb{E}[\widetilde{S}] \leqslant \epsilon_{\widetilde{S}}(\delta)$, if $S_1 > I(X_1; W_1) - H(W_1) + \log p + \delta_{\widetilde{S}}$ and $S_2 > I(X_2; W_2) - H(W_2) + \log p + \delta_{\widetilde{S}}$, where $\epsilon_{\widetilde{S}}, \delta_{\widetilde{S}} \searrow 0$ as $\delta \searrow 0$.*

Proof. The proof is provided in Appendix of [20] □

Now we move on to removing from S the error that is induced due to binning.

Step 2: Error caused by binning

Note that S can be simplified using the definitions of $P^{(\mu_1)}_{M_1|X_1^n}(\cdot|\cdot)$, $P^{(\mu_2)}_{M_2|X_2^n}(\cdot|\cdot)$, and $p^{(\mu)}_{Y^n|\underline{M}}(y^n|\underline{m})$ as

$$S \triangleq \sum_{\underline{x}^n, y^n} \Bigg| p^n_{\underline{X}Y}(\underline{x}^n, y^n) - \sum_{\mu_1,\mu_2} \sum_{\substack{m_1>0, \\ m_2>0}} \sum_{w_1, w_2 \in \mathbb{F}_p^n} \frac{p^n_{\underline{X}}(\underline{x}^n)}{2^{n(C_1+C_2)}} E^{(\mu_1)}_{W_1^n|X_1^n}(w_1^n|x_1^n) E^{(\mu_2)}_{W_2^n|X_2^n}(w_2^n|x_2^n) p^n_{Y|Z}(y^n|f^{(\mu)}(m_1,m_2)) \Bigg|,$$

where $E^{(\mu_i)}_{W_i^n|X_i^n}(w_i^n|x_i^n)$ is defined as $E^{(\mu_i)}_{W_i^n|X_i^n}(w_i^n|x_i^n)$

$$\triangleq p^n \sum_{a_i \in \mathbb{F}_p^k} \frac{p^n_{W_i|X_i}(w_i^n|x_i^n)}{2^{nS_i}(1+\eta)} \mathbb{1}_{\{\mathtt{W}_i(a_i, m_i, \mu_i) = w_1^n\}} \mathbb{1}_{\{w_1^n \in T_\delta(W_1|x_1^n)\}},$$

for $i \in [2]$. Further we bound S using triangle inequality as $S \leqslant S_1 + S_2$, where

$$S_1 \triangleq \sum_{\underline{x}^n, y^n} \Bigg| p^n_{\underline{X}Y}(\underline{x}^n, y^n) - \sum_{\mu_1,\mu_2} \sum_{\substack{m_1>0, \\ m_2>0}} \sum_{w_1^n, w_2^n \in \mathbb{F}_p^n} \frac{p^n_{\underline{X}}(\underline{x}^n)}{2^{n(C_1+C_2)}} E^{(\mu_1)}_{W_1^n|X_1^n}(w_1^n|x_1^n) E^{(\mu_2)}_{W_2^n|X_2^n}(w_2^n|x_2^n) p^{(\mu)}_{Y^n|W^n}(y^n|w_1^n + w_2^n) \Bigg|,$$

$$S_2 \triangleq \sum_{\underline{x}^n, y^n} \sum_{\mu_1,\mu_2} \sum_{\substack{m_1>0, \\ m_2>0}} \sum_{w_1^n, w_2^n \in \mathbb{F}_p^n} \frac{p^n_{\underline{X}}(\underline{x}^n)}{2^{n(C_1+C_2)}} E^{(\mu_1)}_{W_1^n|X_1^n}(w_1^n|x_1^n)$$

$$E^{(\mu_2)}_{W_2^n|X_2^n}(w_2^n|x_2^n)\left|p^n_{Y|Z}(y^n|w_1^n+w_2^n)-p^n_{Y|Z}(y^n|f^{(\mu)}(m_1,m_2))\right|.$$

To bound the term corresponding to S_2, we provide the following proposition.

Proposition 2 (Mutual Packing)**.** *There exist $\epsilon_{S_2}(\delta)$, such that for all sufficiently small δ and sufficiently large n, we have $\mathbb{E}[S_2] \leqslant \epsilon_{S_2}(\delta)$, if $S_1 - R_1 < \log p - H(Z)$, or equivalently, $S_2 - R_2 < \log p - H(Z)$, where $\epsilon_{S_2} \searrow 0$ as $\delta \searrow 0$.*

Proof. The proof is provided in Appendix of [20]. □

Now, we move on to analyzing the term corresponding to S_1.

Step 3: Term concerning Alice's encoding

In this step, we separately analyze the action of the two encoders in approximating the product distribution $p^n_{\underline{X}Y}(\cdot)$. For that, we split S_1 as $S_1 \leqslant Q_1 + Q_2$, where

$$Q_1 \triangleq \sum_{\underline{x}^n, y^n}\left|p^n_{\underline{X}Y}(\underline{x}^n, y^n) - \frac{1}{2^{nC_1}}\sum_{\mu_1}\sum_{m_1>0}\sum_{w_1^n, w_2^n \in \mathbb{F}_p^n} p^n_{\underline{X}}(\underline{x}^n) E^{(\mu_1)}_{W_1^n|X_1^n}(w_1^n|x_1^n) p^n_{W_2^n|X_2^n}(w_2^n|x_2^n) p^n_{Y|Z}(y^n|w_1^n + w_2^n)\right|,$$

$$Q_2 \triangleq \sum_{\underline{x}^n, y^n}\left|\frac{1}{2^{nC_1}}\sum_{\mu_1}\sum_{m_1>0}\sum_{w_1^n, w_2^n \in \mathbb{F}_p^n} p^n_{\underline{X}}(\underline{x}^n) E^{(\mu_1)}_{W_1^n|X_1^n}(w_1^n|x_1^n) \left(p^n_{W_2^n|X_2^n}(w_2^n|x_2^n) - \sum_{m_2>0} E^{(\mu_2)}_{W_2^n|X_2^n}(w_2^n|x_2^n)\right) p^n_{Y|Z}(y^n|w_1^n+w_2^n)\right|.$$

With this partition, the terms within the trace norm of Q_1 differ only in the action of Alice's encoder. And similarly, the terms within the norm of Q_2 differ only in the action of Bob's encoder. Showing that these two terms are small forms a major portion of the achievability proof.

Analysis of Q_1: To prove Q_1 is small, we characterize the rate constraints which ensure that an upper bound to Q_1 can be made to vanish in an expected sense. In addition, this upper bound becomes useful in obtaining a single-letter characterization for the rate needed to make the term corresponding to Q_2 vanish. For this, we define J as

$$J \triangleq \sum_{\underline{x}^n, w_2^n, y^n}\left|p^n_{\underline{X}W_2Y}(\underline{x}^n, y^n) - \frac{1}{2^{nC_1}}\sum_{\mu_1}\sum_{m_1>0}\sum_{w_1^n} p^n_{\underline{X}}(\underline{x}^n) E^{(\mu_1)}_{W_1^n|X_1^n}(w_1^n|x_1^n) p^n_{W_2^n|X_2^n}(w_2^n|x_2^n) p^n_{Y|Z}(y^n|w_1^n + w_2^n)\right|.$$

We now bound the term corresponding to J using the following proposition which is based on the change of measure soft covering discussed in Lemma 1.

Proposition 3. *There exist $\epsilon_J(\delta), \delta_J(\delta)$ such that for all sufficiently small δ and sufficiently large n, we have $\mathbb{E}[J] \leqslant \epsilon_J$ if $S_1 + C_1 \geqslant I(U; RZV)_{\sigma_3} + \log p - H(U)_{\sigma_3} + \delta_J$, where σ_3 is the auxiliary state defined in the theorem and $\epsilon_J, \delta_J \searrow 0$ as $\delta \searrow 0$.*

Proof. The proof is provided in Appendix of [20]. □

Since $Q_1 \leqslant J$, hence Q_1, can be made arbitrarily small for sufficiently large n, if $S_1 + C_1 > I(W_1; X_1X_2YW_2) - H(W_1) + \log p + \delta_J$. Now we move on to bounding Q_2.

Step 4: Analysis of Bob's encoding

Step 3 ensured that the random variables $X_1X_2YW_2$ are close to a product PMF in total variation. In this step, we approximate the PMF of random variables X_1X_2Y using the Bob's encoding rule and bound the therm corresponding to Q_2. We proceed with the following proposition.

Proposition 4. *There exist functions $\epsilon_{Q_2}(\delta)$ and $\delta_{Q_2}(\delta)$, such that for all sufficiently small δ and sufficiently large n, we have $\mathbb{E}[Q_2] \leqslant \epsilon_{Q_2}$, if $S_1 + C_1 \geqslant I(W_1; X_1X_2YW_2) - H(W_1) + \log p + \delta_{Q_2}$ and $S_2 + C_2 \geqslant I(W_2; X_1X_2Y) - H(W_2) + \log p + \delta_{Q_2}$, where $\epsilon_{Q_2}, \delta_{Q_2} \searrow 0$ as $\delta \searrow 0$.*

Proof. The proof is provided in Appendix of [20]. □

Hence, in bounding the terms corresponding to Q_1 and Q_2, we have obtained the following constraints:

$$\begin{aligned} S_1 + C_1 &\geqslant I(W_1; X_1X_2YW_2) - H(W_1) + \log p, \\ S_2 + C_2 &\geqslant I(W_2; X_1X_2Y) - H(W_2) + \log p. \end{aligned} \tag{6}$$

By doing an exact symmetric analysis, but by replacing the first encoder by a product distribution instead of the second encoder in S_1, we obtain the following constraints:

$$\begin{aligned} S_1 + C_1 &\geqslant I(W_1; X_1X_2Y) - H(W_1) + \log p, \\ S_2 + C_2 &\geqslant I(W_2; X_1X_2YW_1) - H(W_2) + \log p. \end{aligned} \tag{7}$$

By time sharing between the above rates (6) and (7), (see [20] for more details) one can obtain the following rate constraints:

$$\begin{aligned} S_1 + C_1 &\geqslant I(W_1; X_1X_2Y) - H(W_1) + \log p, \\ S_2 + C_2 &\geqslant I(W_2; X_1X_2Y) - H(W_2) + \log p, \\ S_1 + S_2 + C_1 + C_2 &\geqslant I(W_1W_2; X_1X_2Y) - H(W_1, W_2) + 2\log p. \end{aligned}$$

D. Rate Constraints

To sum-up, we showed that the (3) holds for sufficiently large n and with probability sufficiently close to 1, if the following bounds holds while incorporating the time sharing random variable Q taking values over the finite set $\mathcal{Q}$[1]:

$$\begin{aligned} S_1 &\geqslant I(X_1; W_1|Q) - H(W_1|Q) + \log p, \\ S_2 &\geqslant I(X_2; W_2|Q) - H(W_2|Q) + \log p, \\ S_1 + C_1 &\geqslant I(X_1X_2Y; W_1|Q) - H(W_1|Q) + \log p, \\ S_2 + C_2 &\geqslant I(X_1X_2Y; W_2|Q) - H(W_2|Q) + \log p, \\ S_1 + S_2 + C_1 + C_2 &\geqslant I(W_1W_2; X_1X_2Y|Q) - H(W_1, W_2|Q) \\ &\qquad + 2\log p, \\ S_1 - R_1 &= S_2 - R_2 \leqslant \log p - H(W_1 \oplus W_2|Q), \\ 0 &\leqslant R_1 \leqslant S_1, \quad 0 \leqslant R_2 \leqslant S_2, \\ C_1 + C_2 &\leqslant C, \quad C \geqslant 0 \end{aligned}$$

Lastly, we complete the proof of the theorem using the Fourier-Motzkin elimination [21] (see [20] for details).

[1]Since Q, the time sharing random variable, is employed in the standard way we omit its discussion here.