Lossy Compression with Near-uniform Encoder Outputs

Badri N. Vellambi, Jörg Kliewer

New Jersey Institute of Technology, Newark, NJ 07102

Email: {badri.vellambi, jkliewer}@njit.edu

Matthieu R. Bloch

Georgia Institute of Technology, Atlanta, GA 30332

Email: matthieu.bloch@ece.gatech.edu

Abstract

It is well known that lossless compression of a discrete memoryless source with near-uniform encoder output is possible at a rate above its entropy if and only if the encoder is randomized. This work focuses on deriving conditions for near-uniform encoder output(s) in the Wyner-Ziv and the distributed lossy compression problems. We show that in the Wyner-Ziv problem, near-uniform encoder output and operation close to the WZ-rate limit is simultaneously possible, whereas in the distributed lossy compression problem, jointly near-uniform outputs is achievable in the interior of the distributed lossy compression rate region if the sources share non-trivial Gács-Körner common information.

Index Terms

Rate-distortion, Slepian-Wolf problem, Wyner-Ziv problem, distributed lossy source coding.

I. Introduction

Owing to the source-channel separation theorem for point-to-point communication and the convenience separation offers, separate source and channel coding and the optimality of separation have been studied in several multi-user problems. Separation-based approaches, especially in multi-user settings, usually assume that the output of source encoders are near-uniform in its alphabet, where uniformity is measured using the variational distance metric. While the lack of near-uniform encoder output(s) does not necessarily cause separation-based approaches to fail, a characterization of when compression of sources can be achieved with near-uniform encoder output(s) simplifies the analysis of separation-based schemes, and is certainly valuable from a theoretical perspective.

Lossless compression with vanishing error probability and near-uniform encoder output was explored in [1], [2]. Hayashi showed that vanishing error probability and near-uniform encoder output cannot be simultaneously achieved [2]. However, one can design lossless codes with near-uniform encoder output if the encoder and decoder share a random seed whose size is roughly the square root of the blocklength of the code [3], [4]. In [4], we have also shown using finite-length results of Kontoyiannis et al. [5] that lossy compression arbitrarily close to the rate-distortion limit is possible even with near-uniform encoder output. In this work, we analyze the rate points for the Wyner-Ziv (WZ) and distributed lossy compression problems at which compression with near-uniform encoder output(s) is possible. Specifically, we have proven the following results.

• *Wyner-Ziv Problem:* Lossy compression with near-uniform encoder output is possible at all rates above the WZ-rate limit.

• *Two-source Distributed Lossy Compression Problem:* If the sources share non-trivial Gács-Körner common information, then lossy compression with jointly near-uniform encoder outputs is achievable at any rate pair in the interior of the distributed lossy compression rate region. The case where the sources share no Gács-Körner common information is open.

The proofs for both problems employ ideas from channel resolvability [6, p. 404] and the likelihood encoder [7]. The result for the distributed lossy compression case is proven without needing a characterization of the underlying rate region. Instead, we exploit the existence of codes with near-uniform encoder outputs for a variant of the Slepian-Wolf problem with a non-standard decoding constraint.

The remainder of the paper is organized thus. Section II provides the notation, Section III defines the problems studied, Section IV details the main results of this work, and lastly, Section V presents the results used in the proofs of Section IV.

II. NOTATION

For $m, n \in \mathbb{N}$ with $m < n$, $[\![m,n]\!] \triangleq \{m, m+1, \ldots, n\}$. Uppercase letters (e.g., X, Y) denote random variables, lower cases denote their realizations (e.g., x, y), and the respective script versions (e.g., $\mathcal{X}$, $\mathcal{Y}$) denote their alphabets. In this work, all alphabets are assumed to be finite. Superscripts indicate the length of vectors, and subscripts indicate the component index. Given a finite set $\mathcal{S}$, $\mathsf{unif}(\mathcal{S})$ denotes the uniform probability mass function (pmf) on $\mathcal{S}$. Given a pmf p_X, $\mathsf{supp}(p_X)$ indicates the support of p_X, $p_X^{\otimes n}$ indicates the joint pmf of n i.i.d random variables distributed according to p_X, and $T_\varepsilon^n[p_X]$ denotes the set of ε-strongly letter typical sequences of length n [8]. Given an event E, $\mathbb{P}(E)$ denotes the probability of its occurrence. Lastly, given two pmfs p and q over a set $\mathcal{X}$, the variational distance is denoted by

$$\mathbb{V}(p, q) \triangleq \sum_{x \in \mathcal{X}} |p(x) - q(x)|. \tag{1}$$

III. PROBLEM DEFINITION

The lossy coding problems studied in this work impose a near-uniform encoder output constraint on the classical Wyner-Ziv and distributed lossy compression problems, and are formally defined here for the sake of completeness.

Definition 1: Let discrete memoryless sources (X, Y) correlated according to pmf Q_{XY}, a bounded distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, d_{\max}]$, and $\Delta \in [0, d_{\max}]$ be given. We say that Wyner-Ziv coding of the source X with receiver side-information Y at an average per-symbol distortion of Δ and is achievable with near-uniform encoder output at a rate $R \in \mathbb{R}^+$ if for every $\varepsilon > 0$, there exist an $n \in \mathbb{N}$, an encoding function $f_X : \mathcal{X}^n \to [\![1, 2^{n(R+\varepsilon)}]\!]$ and a reconstruction function $g_X : [\![1, 2^{n(R+\varepsilon)}]\!] \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n$ at the receiver such that

$$\mathbb{V}(Q_{f_X(X^n)}, \mathsf{unif}([\![1, 2^{n(R+\varepsilon)}]\!])) \leq \varepsilon, \tag{2}$$

$$\textstyle\sum_{i=1}^{n} \mathbb{E}\, d(X_i, \hat{X}_i) \leq n(\Delta + \varepsilon), \tag{3}$$

where $Q_{f_X(X^n)}$ is the pmf of the encoder output $f_X(X^n)$ and $\hat{X}^n = g_X(f_X(X^n), Y^n)$ is the receiver reconstruction.

Definition 2: Let discrete memoryless sources (X, Y) correlated according to pmf Q_{XY}, bounded distortion measures $d_X : \mathcal{X} \times \hat{\mathcal{X}} \to [0, d_{x\max}]$ and $d_Y : \mathcal{Y} \times \hat{\mathcal{Y}} \to [0, d_{y\max}]$, and $\Delta_x \in [0, d_{x\max}]$, $\Delta_y \in [0, d_{y\max}]$ be given. We say that distributed lossy compression with jointly near-uniform encoder outputs and at average per-symbol distortions of Δ_x and Δ_y for sources X and Y, respectively, is achievable at a rate pair $(R_x, R_y) \in \mathbb{R}^{+2}$ if for every $\varepsilon > 0$, there exist an $n \in \mathbb{N}$, encoding functions $f_X : \mathcal{X}^n \to [\![1, 2^{n(R_x+\varepsilon)}]\!]$, $f_Y : \mathcal{Y}^n \to [\![1, 2^{n(R_y+\varepsilon)}]\!]$, and a reconstruction function $g_{XY} : [\![1, 2^{n(R_x+\varepsilon)}]\!] \times [\![1, 2^{n(R_y+\varepsilon)}]\!] \to \hat{\mathcal{X}}^n \times \hat{\mathcal{Y}}^n$ such that

$$\mathbb{V}(Q_{f_X(X^n), f_Y(Y^n)}, Q_U) \leq \varepsilon, \tag{4}$$

$$\textstyle\sum_{i=1}^{n} \mathbb{E}\, d(X_i, \hat{X}_i) \leq n(\Delta_x + \varepsilon), \tag{5}$$

$$\textstyle\sum_{i=1}^{n} \mathbb{E}\, d(Y_i, \hat{Y}_i) \leq n(\Delta_y + \varepsilon), \tag{6}$$

where Q_U is the uniform pmf on $[\![1, 2^{n(R_x+\varepsilon)}]\!] \times [\![1, 2^{n(R_y+\varepsilon)}]\!]$, $Q_{f_X(X^n), f_Y(Y^n)}$ is the pmf of the outputs $f_X(X^n)$, $f_Y(Y^n)$) of the two encoders, and $(\hat{X}^n, \hat{Y}^n) = g_{XY}(f_X(X^n), f_Y(Y^n))$ are the receiver reconstructions.

IV. MAIN RESULTS

A. Near-uniform Wyner-Ziv Coding

Theorem 1: Near-uniform encoder output is achievable in the Wyner-Ziv problem at rates $R \geq \mathsf{R}_{WZ}(\Delta)$.

Proof: The proof builds codes based on channel resolvability [6, p. 404] and the likelihood encoder [7], which allow us to track the distribution of the encoder output more readily than when using the covering lemma. We first pick a channel $Q_{W|X}$ such that:

- the pmf $Q_{W|X}Q_{XY}$ satisfies

$$I(X;W|Y) = I(X;W) - I(Y;W) = \mathsf{R}_{WZ}(\Delta), \tag{7}$$

- $\mathbb{E}[d(X, f(W,Y))] \leq \Delta$ for some function f of (W,Y).

Now, fix $\varepsilon > 0$, and let $R \triangleq I(X;W|Y) + 2\varepsilon$, and $R' \triangleq I(W;Y) - \varepsilon$. Let the codebook $\mathcal{C}$ comprising of $2^{n(R+R')}$ $\hat{X}$-codewords with each codeword selected i.i.d accroding to $Q_W^{\otimes n}$, where Q_W is the marginal of W derived from $Q_{W|X}Q_X$. We arrange the codewords of the random codebook $\mathcal{C}$ in a table of 2^{nR} rows and $2^{nR'}$ columns. Suppose that

$$(K, K') \sim Q_{K,K'} \triangleq \mathsf{unif}(\llbracket 1, 2^{nR} \rrbracket \times \llbracket 1, 2^{nR'} \rrbracket). \tag{8}$$

denotes the random pair of indices used to select the codewords from the codebook $\mathcal{C}$. Let $W^n(K,K')$ be selected and transmitted over the discrete memoryless channel (DMC) $Q_{X|W}$, and $\tilde{X}^n$ be the corresponding output, and let $\tilde{Y}^n$ be the output when $\tilde{X}^n$ is transmitted over the DMC $Q_{Y|X}$.

(a) Channel resolvability code for generating the source X
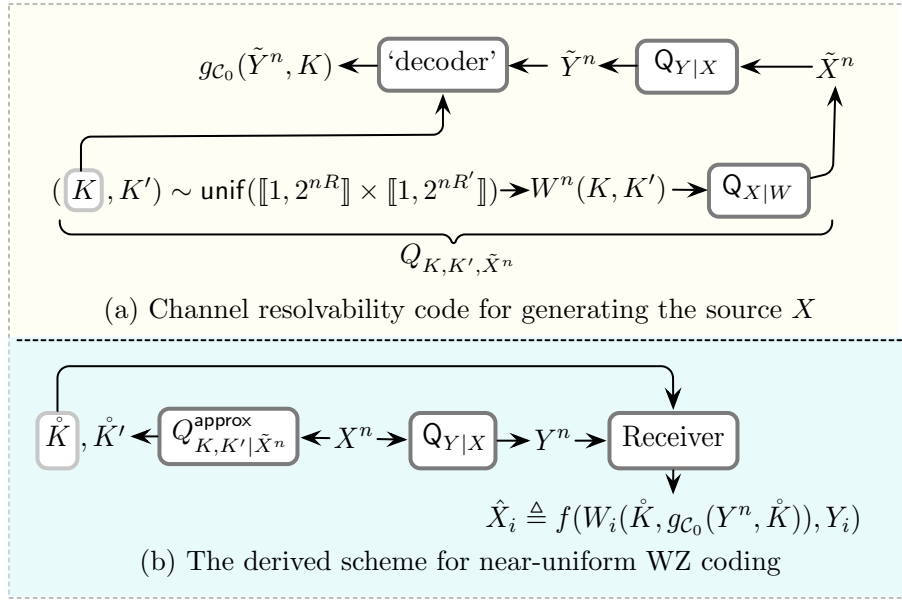
(b) The derived scheme for near-uniform WZ coding

Fig. 1. Source generation and the derived near-uniform WZ scheme.

For this construction, the following hold:
1. Since $R + R' > I(X;W)$, the channel resolvability theorem [6, Theorem 6.3.1] guarantees that

$$\mathbb{E}_{\mathcal{C}}[\mathbb{V}(Q_{\tilde{X}^n}, \mathsf{Q}_X^{\otimes n})] \overset{n \to \infty}{\longrightarrow} 0, \tag{9}$$

where the expectation is over all codebook realizations.
2. Since $R' < I(W;Y)$, there must exist a 'decoding' function $g_{\mathcal{C}} : \mathcal{Y}^n \times \llbracket 1, 2^{nR} \rrbracket \to \llbracket 1, 2^{nR'} \rrbracket$ (depending on $\mathcal{C}$) such that

$$\mathbb{E}_{\mathcal{C}}\left[\mathbb{P}[K' \neq g_{\mathcal{C}}(\tilde{Y}^n, K)]\right] \overset{n \to \infty}{\longrightarrow} 0. \tag{10}$$

3. Since $W^n(K,K') \sim Q_W^{\otimes n}$, and since $W^n(K,K')$ and $(\tilde{X}^n, \tilde{Y}^n)$ are related through the DMC $\mathsf{Q}_{Y|X}Q_{X|W}$, by the weak law of large numbers, we have

$$\mathbb{P}\left[(W^n(K,K'), \tilde{X}^n, \tilde{Y}^n) \notin T_{\varepsilon}^n[Q_{W|X}Q_{XY}]\right] \overset{n \to \infty}{\longrightarrow} 0. \tag{11}$$

Now, for sufficiently large n, we can find a realization $\mathcal{C}_0 = \{w_{\mathcal{C}_0}^n(j,k) : j \in \llbracket 1, 2^{nR} \rrbracket, k \in \llbracket 1, 2^{nR'} \rrbracket\}$ of the codebook such that the sources $\tilde{X}^n$ and $\tilde{Y}^n$ generated by transmitting a codeword selected uniformly at random from $\mathcal{C}_0$ satisfy:

$$\mathbb{V}(Q_{\tilde{X}^n}, \mathsf{Q}_X^{\otimes n}) \leq \varepsilon/2 \tag{12}$$

$$\mathbb{P}[K' \neq g_{\mathcal{C}_0}(\tilde{Y}^n, K)] \leq \varepsilon/2 \tag{13}$$

$$\mathbb{P}\big[(W^n(K, K'), \tilde{X}^n, \tilde{Y}^n) \notin T_\varepsilon^n[Q_{W|X}\mathsf{Q}_{XY}]\big] \leq \varepsilon/2. \tag{14}$$

Let $Q_{K,K'\tilde{X}^n}$ be the pmf induced by the codebook $\mathcal{C}_0$. Now, to derive a (randomized) WZ scheme from this channel resolvability code, we proceed as given in Fig. 1. We first pick an approximation $Q_{K,K',\tilde{X}^n}^{\text{approx}}$ of $Q_{K,K',\tilde{X}^n}$ such that

$$\mathbb{V}(Q_{K,K',\tilde{X}^n}^{\text{approx}}, Q_{K,K',\tilde{X}^n}) \leq \varepsilon/2. \tag{15}$$

The need for an approximation will become clear later when we *emulate* $Q_{K,K',\tilde{X}^n}$ using $\tilde{X}^n$ and a near-uniform random seed. Upon choosing $Q_{K,K',\tilde{X}^n}^{\text{approx}}$, we encode X^n by generating $(\mathring{K}, \mathring{K}') \sim Q_{K,K'|\tilde{X}^n}^{\text{approx}}(\cdot, \cdot|X^n)$. We then transmit only $\mathring{K}$ to the receiver. The joint pmf of $(\mathring{K}, \mathring{K}', X^n)$ is given by

$$Q_{\mathring{K},\mathring{K}',X^n}(\mathring{k}, \mathring{k}', x^n) \triangleq Q_{K,K'|\tilde{X}^n}^{\text{approx}}(\mathring{k}, \mathring{k}'|x^n)\mathsf{Q}^{\otimes n}(x^n). \tag{16}$$

From (12), (15) and (16), we are guaranteed that

$$\mathbb{V}(Q_{\mathring{K},\mathring{K}',X^n}, Q_{K,K',\tilde{X}^n}) \leq \mathbb{V}(Q_{\mathring{K},\mathring{K}',\tilde{X}^n}^{\text{approx}}, Q_{K,K',\tilde{X}^n}) + \mathbb{V}(Q_{\tilde{X}^n}, \mathsf{Q}_X^{\otimes n}) \leq \varepsilon. \tag{17}$$

Further, since Y^n and $\tilde{Y}^n$ are the outputs when X^n and $\tilde{X}^n$, respectively, are fed into the DMC $\mathsf{Q}_{Y|X}$, we are guaranteed to have

$$\mathbb{V}(Q_{\mathring{K},\mathring{K}',X^n,Y^n}, Q_{K,K',\tilde{X}^n,\tilde{Y}^n}) \leq \varepsilon. \tag{18}$$

Consequently, the following also hold

$$\mathbb{V}(Q_{\mathring{K},\mathring{K}',Y^n}, Q_{K,K',\tilde{Y}^n}) \leq \varepsilon \tag{19}$$

$$\mathbb{V}(Q_{\mathring{K},\mathring{K}'}, Q_{K,K'}) \leq \varepsilon, \tag{20}$$

From (8) and (20), we see that $\mathring{K}$ and $\mathring{K}'$ are jointly nearly uniform. Hence, $\mathring{K}$, which is the WZ encoder output, is also nearly uniform. Further, (14) and (18) jointly imply that

$$\mathbb{P}\big[(W^n(\mathring{K}, \mathring{K}'), X^n, Y^n) \notin T_\varepsilon^n[Q_{W|X}\mathsf{Q}_{XY}]\big] \leq 3\varepsilon/2. \tag{21}$$

Next, from (13), (18) and Lemma 1 of Section V, we see that:

$$\mathbb{P}[\mathring{K}' \neq g_{\mathcal{C}_0}(Y^n, \mathring{K})] \leq 3\varepsilon/2, \tag{22}$$

$$\mathbb{P}\big[W^n(\mathring{K}, \mathring{K}') \neq W^n(\mathring{K}, g_{\mathcal{C}_0}(Y^n, \mathring{K}))\big] \leq 3\varepsilon/2. \tag{23}$$

Combining (21) and (23), we conclude that

$$\mathbb{P}\big[(W^n(\mathring{K}, g_{\mathcal{C}_0}(Y^n, \mathring{K})), X^n, Y^n) \notin T_\varepsilon^n[Q_{W|X}\mathsf{Q}_{XY}]\big] \leq 3\varepsilon. \tag{24}$$

Thus, if the receiver estimates $\mathring{K}'$ using $g_{\mathcal{C}_0}(Y^n, \mathring{K})$, and sets $\hat{X}_i \triangleq f(W_i(\mathring{K}, g_{\mathcal{C}_0}(Y^n, \mathring{K})), Y_i)$ as the reconstruction for X_i, $i = 1, \ldots, n$, then with a probability of $1 - 3\varepsilon$, the per-symbol distortion is at most $\Delta(1 + 3\varepsilon)$, since

$$\mathbb{P}\Big[d(X^n, f(W^n(\mathring{K}, \mathring{g}_{\mathcal{C}_0}(Y^n, \mathring{K})), Y^n)) > \Delta(1+\varepsilon)\Big] \leq 3\varepsilon. \tag{25}$$

Thus, we are guaranteed to have an average per-symbol distortion of no more than $\Delta + 3\varepsilon d_{\max}$. We are nearly done, if we ensure that:

(1) a suitable $Q_{K,K'\tilde{X}^n}^{\text{approx}}$ is selected; and

(2) the encoding is deterministic. (The encoding above involves randomly generating $(\mathring{K}, \mathring{K}')$ using $Q_{K,K'\tilde{X}^n}^{\text{approx}}$.)

We can guarantee the first requirement by invoking Lemma 2 of Section V, which ensures that an approximation $Q_{K,K',\tilde{X}^n}^{\text{approx}}$ of $Q_{K,K',\tilde{X}^n}$ meeting (15) can be realized if the encoder is given a uniform random seed of rate $R + R' - I(X; W) + \varepsilon = 2\varepsilon$ that is independent of $\tilde{X}^n$. We can ensure the second requirement by approximating this uniform seed by a near-uniform seed of rate 2ε obtained as a function of $\{X_{n+\ell} : \ell = 1, \ldots, \frac{3\varepsilon n}{H(X)}\}$ that extracts its intrinsic randomness (Lemma 6 of Section V). Thus, both the pmf $Q_{\mathring{K},\mathring{K}',X^n}$ of (16) and the encoding

operation can be realized as a deterministic function of $n + \frac{3\varepsilon n}{H(X)}$ symbols of the X source.

Finally, since the last $\frac{3\varepsilon n}{H(X)}$ source symbols are used solely to generate the random seed, it can be assumed that the average distortion corresponding to each of these symbols is no more than $d_{\max}$. Combining this with the estimate for the first n symbols, we see that the overall average per-symbol distortion offered by the code is at most $\Delta + 3\varepsilon d_{\max} + \frac{3\varepsilon}{H(X)} d_{\max}$. The result then follows by limiting ε to zero. ∎

B. Near-uniform Distributed Lossy Source Coding Problem

We begin by analyzing joint near-uniformity of encoder outputs in a variant of the Slepian-Wolf (SW) problem, which will be used for the corresponding distributed lossy compression problem. Since the lossless compression of a source with near-uniform output is not possible without shared randomness between encoder and decoder [4], SW coding with jointly near-uniform encoder outputs is also not possible. However, if we relax the decoder constraint to lossless recovery of all but a small fraction of symbols, then there exist distributed coding schemes with jointly near-uniform encoder outputs provided the two sources share non-trivial Gács-Körner common information. The following result quantifies this precisely.

Theorem 2: Let $(X, Y) \sim Q_{XY}$ and suppose that the random variable U common to X and Y (in the Gács-Körner sense) be non-trivial. Let $\varepsilon \in (0, H(U))$. Then, for any (R_x, R_y) in the interior of the Slepian-Wolf rate region, there exist $n \in \mathbb{N}$ and $m \in [\![n, n + \frac{9\varepsilon n}{H(U)}]\!]$, encoding functions $f_X : \mathcal{X}^m \to [\![1, 2^{nR_x}]\!]$ and $f_Y : \mathcal{Y}^m \to [\![1, 2^{nR_y}]\!]$ operating over m source symbols, and a decoding function $g_{XY} : [\![1, 2^{nR_x}]\!] \times [\![1, 2^{nR_y}]\!] \to \mathcal{X}^n \times \mathcal{Y}^n$ outputting n symbols of both sources such that

$$\mathbb{V}(Q_{f_X(X^m), f_Y(Y^m)}, \mathsf{unif}([\![1, 2^{nR_x}]\!] \times [\![1, 2^{nR_y}]\!])) \leq \varepsilon, \tag{26}$$

$$\mathbb{P}[(X^n, Y^n) \neq g_{XY}(f_X(X^m), f_Y(Y^m))] \leq \varepsilon. \tag{27}$$

Proof: The proof proves that the claim holds for a corner point of the SW rate region, which extends to the other corner point by reversing the roles of the sources, and to the interior of the rate region by time-sharing. Without loss of generality, let us build a coding scheme for the corner point at which Y is available at the decoder. Let U indicate the Gács-Körner common randomness between X and Y. Let

$$[R_u \ R_x \ R_y] \triangleq [H(U) + \varepsilon/2 \ \ H(X|Y) + \varepsilon/2 \ \ H(Y|U) + \varepsilon/2] \tag{28}$$

$$R_x' \triangleq I(X; Y|U) + \varepsilon/2 < I(X; Y) \tag{29}$$

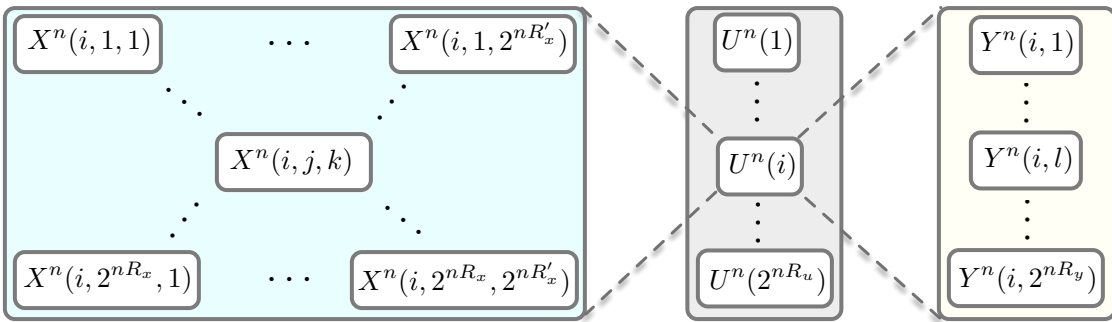Thus, $R_u + R_y > H(U, Y) = H(Y)$, and $R_x > H(X|Y)$.

Fig. 2.   Codebook setup for the Slepian-Wolf problem

As illustrated in Fig. 2, a random codebook $\{U^n(1), \ldots, U^n(2^{nR_u})\}$ by choosing codewords i.i.d. according to $Q_U^{\otimes n}$. For each $i \in [\![1, 2^{nR_U}]\!]$, generate a codebook of X-codewords arranged as 2^{nR_x} rows and $2^{nR_x'}$ columns with codewords selected i.i.d. using $Q_{X|U}^{\otimes n}(\cdot|U^n(i))$. Note that this codebook has $2^{n(H(X|U)+\varepsilon)}$ entries. Next, for each $i \in [\![1, 2^{nR_U}]\!]$, generate a codebook of 2^{nR_y} Y-codewords with codewords selected i.i.d. using $Q_{Y|U}^{\otimes n}(\cdot|U^n(i))$. We let $\mathcal{C}$ to jointly represent the three codebooks. Now, let random indices I, J, K, L satisfy

$$Q_{I,J,K} \triangleq \mathsf{unif}([\![1, 2^{nR_u}]\!] \times [\![1, 2^{nR_x}]\!] \times [\![1, 2^{nR_x'}]\!]), \tag{30}$$

$$Q_{I,L} \triangleq \mathsf{unif}([\![1, 2^{nR_u}]\!] \times [\![1, 2^{nR_y}]\!]). \tag{31}$$

Let $\hat{U}^n \triangleq U^n(I)$, $\hat{X}^n \triangleq X^n(I, J, K)$ and $\tilde{Y}^n \triangleq Y^n(I, L)$, and let $\hat{Y}^n$ be the output of the DMC $Q_{Y|X}$ when the input is $\hat{X}^n = X^n(I, J, K)$. By an application of Lemma 5 of Section V, we see that

$$\mathbb{E}_{\mathcal{C}}[D_{KL}(Q_{\hat{X}^n} \| Q_X^{\otimes n})] \overset{n \to \infty}{\longrightarrow} 0, \tag{32}$$

$$\mathbb{E}_{\mathcal{C}}[D_{KL}(Q_{\tilde{Y}^n} \| Q_Y^{\otimes n})] \overset{n \to \infty}{\longrightarrow} 0, \tag{33}$$

since $R_u + R_x + R_x' > H(U, X) = H(X)$, $R_u + R_y > H(Y, U) = H(Y)$, and $R_u > H(U)$. Note that since $\hat{Y}^n$ and Y^n are obtained by transmitting $\hat{X}^n$ and X^n, respectively, on the DMS $Q_{Y|X}$, we are also guaranteed that

$$\mathbb{E}_{\mathcal{C}}[D_{KL}(Q_{\hat{X}^n \hat{Y}^n} \| Q_{XY}^{\otimes n})] \overset{n \to \infty}{\longrightarrow} 0, \tag{34}$$

Further, using a similar argument, we can also show that

$$\mathbb{E}_{\mathcal{C}}\left[2^{-nR_u} \sum_i D_{KL}\big(Q_{\hat{Y}^n|I=i} \| Q_{Y|U}^{\otimes n}(\cdot|U^n(i))\big) \right] \overset{n \to \infty}{\longrightarrow} 0, \tag{35}$$

$$\mathbb{E}_{\mathcal{C}}\left[2^{-nR_u} \sum_i D_{KL}\big(Q_{\tilde{Y}^n|I=i} \| Q_{Y|U}^{\otimes n}(\cdot|U^n(i))\big) \right] \overset{n \to \infty}{\longrightarrow} 0. \tag{36}$$

Now, let $\eta_n \triangleq \mathbb{E}[D_{KL}(Q_{\hat{Y}^n, I, J} \| Q_{\hat{Y}^n, I} Q_J)]$. Then, the following argument holds.

$$\eta_n = \mathbb{E}_{\mathcal{C}}\left[D_{KL}(Q_{\hat{Y}^n, I, J} \| Q_{\hat{Y}^n, I} Q_J) \right] \tag{37}$$

$$= \mathbb{E}_{\mathcal{C}}\left[D_{KL}(Q_{\hat{Y}^n, I, J} \| Q_I Q_J Q_{Y|U}^{\otimes n}(\cdot|U^n(I))) - D_{KL}(Q_{\hat{Y}^n, I} \| Q_I Q_{Y|U}^{\otimes n}(\cdot|U^n(I))) \right] \tag{38}$$

$$\overset{(a)}{\leq} \mathbb{E}_{\mathcal{C}}\left[D_{KL}(Q_{\hat{Y}^n, I, J} \| Q_{I,J} Q_{Y|U}^{\otimes n}(\cdot|U^n(I))) \right] \tag{39}$$

$$\overset{(b)}{=} \mathbb{E}_{\mathcal{C}}\left[\sum_{y^n} \left[\frac{\sum_k Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k))}{2^{nR_x'}} \right] \log_2 \frac{\sum_{k'} Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k'))}{2^{nR_x'} Q_{Y|U}^{\otimes n}(y^n|U^n(1))} \right] \tag{40}$$

$$= \sum_{y^n, k} \mathbb{E}_{U^n(1), X^n(1,1,k)} \left[\frac{Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k))}{2^{nR_x'}} \mathbb{E}_{\text{rest}} \left[\log_2 \frac{\sum_{k'} Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k'))}{2^{nR_x'} Q_{Y|U}^{\otimes n}(y^n|U^n(1))} \bigg|_{X^n(1,1,k)}^{U^n(1)} \right] \right] \tag{41}$$

$$\leq \sum_{y^n, k} \mathbb{E}_{U^n(1), X^n(1,1,k)} \left[\frac{Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k))}{2^{nR_x'}} \log_2 \mathbb{E}_{\text{rest}} \left[\frac{\sum_{k'} Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k'))}{2^{nR_x'} Q_{Y|U}^{\otimes n}(y^n|U^n(1))} \bigg|_{X^n(1,1,k)}^{U^n(1)} \right] \right] \tag{42}$$

$$\leq \sum_{y^n, k} \mathbb{E}_{U^n(1), X^n(1,1,k)} \left[\frac{Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k))}{2^{nR_x'}} \log_2 \left[1 + \frac{Q_{Y|X}^{\otimes n}(y^n|X^n(1,1,k))}{2^{nR_x'} Q_{Y|U}^{\otimes n}(y^n|U^n(1))} \right] \right] \tag{43}$$

$$\leq \log_2\left(1 + 2^{n(I(X;Y|U) - R_x' + 2\delta \log_2 |\mathcal{Y}|)}\right) + 2|\mathcal{X}||\mathcal{Y}||\mathcal{U}|e^{-n\delta^2 \mu} \log_2\left(1 + \mu^{-n}\right) \overset{n \to \infty}{\longrightarrow} 0 \quad \text{(due to (29))}, \tag{44}$$

where

- (a) follows by dropping the second non-negative term that is subtracted;
- (b) due to the i.i.d. construction of the random codebooks;
- (41) uses the law of iterated expectations, where $\mathbb{E}_{\text{rest}}$ is the expectation over all codewords except $(U^n(1), X^n(1,1,k))$;
- (42) uses Jensen's inequality for the $\log$ function;
- (43) because $X^n(1,1,k') \sim Q_{X|U}^{\otimes n}(\cdot|U^n(1))$ for $k' \neq k$, and

$$\mathbb{E}_{\text{rest}}\left[Q_{Y|X}^{\otimes n}(\cdot|X^n(1,1,k')) \big| X^n(1,1,k) \right] = Q_{Y|U}^{\otimes n}(\cdot|U^n(1)), \tag{45}$$

  since $X^n(1,1,k')$ is chosen using $\prod_{\ell=1}^n Q_{X|U}(\cdot|U_\ell(1))$; and

- finally, (44) follows by splitting the outer sum depending on whether the realization of the codeword $X^n(1,1,k)$ and y^n are jointly δ-strongly letter typical, where $\delta < \varepsilon/(4\log_2 |\mathcal{Y}|)$, and

$$\mu \triangleq \min\{Q_{X,Y,U}(x, y, u) : (x, y, u) \in \text{supp}(Q_{X,Y,U})\}. \tag{46}$$

Note that because of the choice of R'_x in (29), the bound in (44) approaches 0 as $n \to \infty$. From (32)-(36) and (44), we conclude that there must exist for a sufficiently large n, a codebook $\mathcal{C}^*$ such that

$$\mathbb{V}(Q_{\hat{U}^n}, \mathsf{Q}_U^{\otimes n}) < \varepsilon \tag{47}$$

$$\mathbb{V}(Q_{\hat{X}^n \hat{Y}^n}, \mathsf{Q}_{XY}^{\otimes n}) < \varepsilon, \tag{48}$$

$$\mathbb{V}(Q_{\tilde{Y}^n}, \mathsf{Q}_Y^{\otimes n}) < \varepsilon, \tag{49}$$

$$\mathbb{V}(Q_{\hat{Y}^n, I, J}, Q_{\hat{Y}^n, I} Q_J) < \varepsilon, \tag{50}$$

$$\mathbb{V}(Q_{\hat{Y}^n, I}, Q_{\tilde{Y}^n, I}) < \varepsilon. \tag{51}$$

The code $\mathcal{C}^*$ induces two joint pmfs $Q^*_{I,J,K,\hat{U}^n,\hat{X}^n,\hat{Y}^n}$ and $Q^*_{I,L,\hat{U}^n \tilde{Y}^n}$ for which $Q^*_{I,J,K} = Q_{I,J,K}$ and $Q^*_{I,L} = Q_{I,L}$. Since (51) holds, there must exist a joint pmf $Q^\dagger_{\hat{Y}^n, \tilde{Y}^n, I}$ over $\mathcal{Y}^n \times \mathcal{Y}^n \times [\![1, 2^{nR_u}]\!]$ that optimally couples $(\hat{Y}^n, I)$ and $(\tilde{Y}^n, I)$ so that

$$\mathbb{P}[(\hat{Y}^n, I) \neq (\tilde{Y}^n, I)] \leq 2\varepsilon. \tag{52}$$

Further, since $\hat{Y}^n$ and $\tilde{Y}^n$ are generated from the same U-codebook, they share $\hat{U}^n$ as common randomness in the Gács-Körner sense. Now, let $Q^\circ_{I,J,K,\hat{X}^n,\tilde{Y}^n,L}$ be the marginal pmf of $(I, J, K, \hat{X}^n, \tilde{Y}^n, L)$ obtained from

$$Q^\circ_{I,J,K,\hat{X}^n,\hat{Y}^n,\tilde{Y}^n,L} \triangleq Q^*_{I,J,K,\hat{X}^n,\hat{Y}^n} Q^\dagger_{\tilde{Y}^n|\hat{Y}^n,I} Q^*_{L|I,\tilde{Y}^n}. \tag{53}$$

For the pmf $Q^\circ_{I,J,K,\hat{X}^n,\tilde{Y}^n,L}$, we can show the following:

$$\mathbb{V}(\mathsf{Q}_{XY}^{\otimes n}, Q^\circ_{\hat{X}^n \tilde{Y}^n}) \overset{(48),(52),(53)}{\leq} 3\varepsilon, \tag{54}$$

$$\mathbb{V}(Q^\circ_{\tilde{Y}^n, I, J}, Q^\circ_{\tilde{Y}^n, I} Q_J) \overset{(50),(52),(53)}{\leq} 5\varepsilon, \tag{55}$$

$$\mathbb{V}(Q^\circ_{\tilde{Y}^n, I, L}, Q^*_{\tilde{Y}^n, I, L}) \overset{(52)}{\leq} 2\varepsilon, \tag{56}$$

Note that even though $Q^\circ_{\tilde{Y}^n, I, L}$ and $Q^*_{\tilde{Y}^n, I, L}$ could be different, we can still view $\tilde{Y}^n$ as being generated using the two-stage codebook by first choosing the U-codeword uniformly at random, and then the Y-codeword by selecting the index L according to $Q^\circ_{L|I}$, which is only nearly uniform. Lastly, since in Q°, we have $L-(I, \tilde{Y}^n)-(J, K)$, we also have

$$\mathbb{V}(Q^\circ_{L,I,J}, Q^\circ_{L,I} Q_J) \overset{(55),(53)}{\leq} 5\varepsilon, \tag{57}$$

$$\mathbb{V}(Q^\circ_{L,I,J}, Q^*_{L,I} Q_J) \overset{(57),(56)}{\leq} 7\varepsilon. \tag{58}$$

Thus, under the law Q°, (I, L) and J are jointly nearly-uniform. We now use an approach similar to the Wyner-Ziv case to build a code for the problem at hand.

- The X-encoder first generates $I^\circ \sim Q^\circ_{I|\hat{U}^n}(\cdot|U^n)$, and then $(J^\circ, K^\circ) \sim Q^\circ_{J,K|\hat{X}^n,I}(\cdot|X^n, I^\circ)$. It sends J° to the receiver;
- The Y-encoder generates $I^\circ \sim Q^\circ_{I|\hat{U}^n}(\cdot|U^n)$ that matches the index generated by the X-encoder, and then generates $L^\circ \sim Q^\circ_{L|\tilde{Y}^n,I}(\cdot|Y^n, I^\circ)$. It sends (I°, L°) to the receiver;
- The decoder declares $Y^n(I^\circ, L^\circ)$ as the realization of Y^n. It then looks for an index K such that $X^n(I^\circ, J^\circ, K)$ is jointly typical with $Y^n(I^\circ, L^\circ)$. With high probability, the search will yield a unique K that matches K°, since $K \in [\![1, 2^{nR'_x}]\!]$ and $R'_x < I(X;Y)$ (see (29)).

The above encoding and decoding operations emulate the following joint pmf of sources and indices:

$$\mathsf{Q}_{XY}^{\otimes n} Q^\circ_{I|\hat{U}^n}(\cdot|U^n) Q^\circ_{J,K|\hat{X}^n,I}(\cdot|X^n, \cdot) Q^\circ_{L|\tilde{Y}^n,I}(\cdot|Y^n, \cdot). \tag{59}$$

According to (54), the variational distance between the emulated pmf and $Q^\circ_{I,J,K,\hat{X},\tilde{Y},L}$ is no more than 3ε, which when combined with (58) implies that the variational distance of the emulated joint pmf of $(I^\circ, J^\circ, L^\circ)$ is at most 10ε away from the jointly uniform pmf $Q_{L,I} Q_J$. Lastly, as in the Wyner-Ziv case, we are done if we approximate

the randomized encoders by functions, for which we use near-uniform seeds derived from additional source symbols in the following manner.

- At both encoders, we use $U_{n+1}, \ldots, U_{n+\frac{3\varepsilon n}{H(U)}}$ to obtain the same near-uniform random seed over $[\![1, 2^{2n\varepsilon}]\!]$, and then use the seed to approximate the random index selection according to $Q^{\circ}_{I|\hat{U}^n}(\cdot|U^n)$.

- We use $X_{n+\frac{3\varepsilon n}{H(U)}+1}, \ldots, X_{n+\frac{6\varepsilon n}{H(U)}}$ to obtain a near-uniform random seed over $[\![1, 2^{2n\varepsilon}]\!]$, and then use the seed to realize index selection according to $Q^{\circ}_{J,K|\hat{X}^n,I}(\cdot|X^n, \cdot)$.

- We use $Y_{n+\frac{6\varepsilon n}{H(U)}+1}, \ldots, X_{n+\frac{9\varepsilon n}{H(U)}}$ to obtain a near-uniform random seed over $[\![1, 2^{2n\varepsilon}]\!]$, and then use the seed to realize index selection according to $Q^{\circ}_{L|\tilde{Y}^n,I}(\cdot|Y^n, \cdot)$.

In the above, extracting random seeds and realizing index selections as a function of the random seed and the sources are done by invoking Lemmas 2 and 6 of Section V.

Thus, for sufficiently large n, there exist codes that encode $n + \frac{9\varepsilon n}{H(U)}$ source symbols into a jointly nearly uniformly distributed pair of indices, using which the first n symbols can be losslessly retrieved with high probability. ∎

We are now ready to present our result pertaining to uniform lossy compression in the two-source distributed lossy source coding problem. Note that the proof does not require a characterization of the underlying rate region.

Theorem 3: Given jointly correlated sources $(X, Y) \sim Q_{XY}$ with non-trivial Gács-Körner common information, distributed lossy compression with jointly near-uniform encoder outputs is possible at all rate points in the strict interior of the distributed lossy compression rate region.

Proof: Let (R_x, R_y) be in the interior of the distributed lossy compression rate region. Fix $\varepsilon > 0$. Then, for sufficiently large n, there exist encoders f_X and f_Y operating at rates no more than $R_x + \varepsilon$ and $R_y + \varepsilon$, and a reconstruction function g_{XY} that operates on the encoder outputs to generate reconstructions for X and Y with an average per-symbol distortion of at most $\Delta_x + \varepsilon$ and $\Delta_y + \varepsilon$, respectively. Without loss of generality, we may assume that $f_X(X^n)$ and $f_Y(Y^n)$ share non-trivial Gács-Körner common information. Else, we can increase the encoded message rates by ε by appending to each encoder output, a function of U – the random variable common to X and Y in the Gács-Körner sense.

Now, let $\mathring{X} = f_X(X^n)$ and $\mathring{Y} = f_Y(Y^n)$, and let $\mathring{U}$ be the random variable common to $\mathring{X}$ and $\mathring{Y}$ in the Gács-Körner sense. From Theorem 2, we see that there exists sufficiently large $N \in \mathbb{N}$, sufficiently small δ, and $M \leq N + 9\delta N$ such that there exists a code that encodes M symbols of the correlated source $(\mathring{X}, \mathring{Y})$ in any interior point of its SW rate region and recovers the first N source symbols of $\mathring{X}$ and $\mathring{Y}$ losslessly. Concatenating M copies of the lossy source code with encoders f_X and f_Y (as the outer code) followed by the above code for $\mathring{X}$ and $\mathring{Y}$ (as the inner code) will yield a joint code operating at rates of no more than $R_x + 2\varepsilon + \delta$ and $R_y + 2\varepsilon + \delta$, respectively. Moreover, the average distortions offered by this joint code for the nM symbols of X and Y are at most $\frac{\Delta_x + \varepsilon + 9\delta d_{x\max}}{1+9\delta}$ and $\frac{\Delta_y + \varepsilon + 9\delta d_{y\max}}{1+9\delta}$, respectively. Since ε and δ are arbitrary, the claim holds. ∎

V. REQUIRED RESULTS

Lemma 1: Let p.m.f. $Q_{A,B}$ over a finite set $\mathcal{A} \times \mathcal{B}$ be such that for $(A, B) \sim Q_{AB}$, there exists a function $\phi(B)$ such that $\mathbb{P}[A \neq \phi(B)] \leq \varepsilon$. Now, let $(\tilde{A}, \tilde{B}) \sim \tilde{Q}_{\tilde{A},\tilde{B}}$ be such that $\mathbb{V}(\tilde{Q}_{\tilde{A},\tilde{B}}, Q_{A,B}) \leq \varepsilon$. Then, $\mathbb{P}[\tilde{A} \neq \phi(\tilde{B})] \leq 2\varepsilon$.

Proof: Let $\mathcal{S} = \{(a, b) : a \neq \phi(b)\}$. Then,

$$\mathbb{P}[A \neq \phi(B)] = Q_{A,B}(\mathcal{S}) \tag{60}$$

$$\mathbb{P}[\tilde{A} \neq \phi(\tilde{B})] = \tilde{Q}_{\tilde{A},\tilde{B}}(\mathcal{S}). \tag{61}$$

Thus,

$$|\mathbb{P}[\tilde{A} \neq \phi(\tilde{B})] \leq \mathbb{P}[A \neq \phi(B)]| + |\tilde{Q}_{\tilde{A},\tilde{B}}(\mathcal{S}) - Q_{A,B}(\mathcal{S})|$$

$$\leq \varepsilon + \sum_{(a,b) \in \mathcal{S}} |\tilde{Q}_{\tilde{A},\tilde{B}}(a, b) - Q_{A,B}(a, b)| \tag{62}$$

$$\leq \varepsilon + \mathbb{V}(\tilde{Q}_{\tilde{A},\tilde{B}}, Q_{A,B}) \leq 2\varepsilon. \tag{63}$$

∎

Lemma 2: Given p.m.f. Q_{AB} over a finite alphabet $\mathcal{A} \times \mathcal{B}$ and $R > I(A; B)$, suppose that we construct a random codebook $\mathcal{C}_n$ of 2^{nR} A-codewords generated randomly using Q_A. Let $L \sim \text{unif}([\![1, 2^{nR}]\!])$. Suppose that

$A^n(L)$ is transmitted over the DMC $Q_{B|A}$ and $\tilde{B}^n$ is the corresponding output. Let $S \sim \mathsf{unif}(\llbracket 1, 2^{n\rho} \rrbracket)$, where $\rho > R - I(A;B)$. Then, there exists $\phi_{\mathcal{C}_n} : \mathcal{B}^n \times \llbracket 1, 2^{n\rho} \rrbracket \to \llbracket 1, 2^{nR} \rrbracket$ (that depends on $\mathcal{C}_n$) such that

$$\lim_{n \to \infty} \mathbb{E}\left[\mathbb{V}(Q_{\phi(\tilde{B}^n, S), \tilde{B}^n}, Q_{L, \tilde{B}^n}) \right] = 0, \tag{64}$$

where $Q_{L, \tilde{B}^n}$ is the joint p.m.f. of $(L, \tilde{B}^n)$ induced by $\mathcal{C}_n$.

Proof: Let $\delta, \varepsilon > 0$ be chosen such that

$$\rho - R + I(A;B) - 4\delta \log_2(|\mathcal{A}||\mathcal{B}|) > \varepsilon. \tag{65}$$

By the random codebook construction, it follows that $(A^n(L), \tilde{B}^n)$ is as if it is the output from a DMS Q_{AB}. Hence, by [8, Theorem 1.1], it follows that

$$\mathbb{P}\left[(A^n(L), \tilde{B}^n) \notin T_\delta^n[Q_{AB}] \right] \leq 2Me^{-n\delta^2\mu}, \tag{66}$$

where $M \triangleq |\mathcal{A}||\mathcal{B}|$ and $\mu \triangleq \min_{(a,b) \in \mathsf{supp}(Q_{AB})} Q_{AB}(a,b)$. Now, let for a codebook $\mathcal{C}_\circ \triangleq \{(a^n(l)\}_{l \in \llbracket 1, 2^{nR} \rrbracket}$ and $b^n \in \mathcal{B}^n$,

$$\mathcal{L}_{\mathcal{C}_\circ}(b^n) \triangleq \{l : (a^n(l), b^n) \in T_\delta^n[Q_{AB}]\}. \tag{67}$$

From Lemma 3 below, the following holds for sufficiently large n.

$$\mathbb{E}\left[|\mathcal{L}_{\mathcal{C}}(\tilde{B}^n)| \right] \leq 2^{1+n(R-I(A;B)+2\delta \log_2 M)}. \tag{68}$$

Let $\mathcal{F}$ be the collection of codebooks $\mathcal{C}_0 \triangleq \{a^n(l)\}_{l \in \llbracket 1, 2^{nR} \rrbracket}$ such that the following hold.

$$\mathbb{P}\left[(a^n(L), \tilde{B}^n) \notin T_\delta^n[Q_{AB}] \,\middle|\, \mathcal{C} = \mathcal{C}_0 \right] \leq \sqrt{2Me^{-n\delta^2\mu}} \tag{69}$$

$$\frac{\mathbb{E}\left[|\mathcal{L}_{\mathcal{C}}(\tilde{B}^n)| \,\middle|\, \mathcal{C} = \mathcal{C}_0 \right]}{\mathbb{E}\left[|\mathcal{L}_{\mathcal{C}}(\tilde{B}^n)| \right]} \leq 2^{\delta \log_2 M}. \tag{70}$$

By Markov's inequality, we then have

$$\mathbb{P}[\mathcal{C} \notin \mathcal{F}] \leq \sqrt{2Me^{-n\delta^2\mu}} + 2^{-n\delta \log_2 M}. \tag{71}$$

Now, pick $\mathcal{C}^* \triangleq \{a^{*n}(l)\}_{l \in \llbracket 1, 2^{nR} \rrbracket} \in \mathcal{F}$ and define $\mathcal{G}_{\mathcal{C}^*}$ as the set of all b^n such that

$$\mathbb{P}\left[(a^{*n}(L), \tilde{B}^n) \notin T_\delta^n[Q_{AB}] \,\middle|\, \begin{matrix} \tilde{B}^n = b^n \\ \mathcal{C} = \mathcal{C}^* \end{matrix} \right] \leq \sqrt[4]{2Me^{-n\delta^2\mu}} \tag{72}$$

$$|\mathcal{L}_{\mathcal{C}^*}(b^n)| \leq 2^{1+n(R-I(A;B)+4\delta \log_2 S)}. \tag{73}$$

Again, by Markov's inequality, it follows that

$$\mathbb{P}[\tilde{B}^n \notin \mathcal{G}_{\mathcal{C}^*} \mid \mathcal{C} = \mathcal{C}^*] \leq \eta_0 \triangleq \sqrt[4]{2Me^{-n\delta^2\mu}} + 2^{-n\delta \log_2 M}.$$

Further, it also follows that for each $b^n \in \mathcal{G}_{\mathcal{C}^*}$,

$$\sum_{l \notin \mathcal{L}_{\mathcal{C}^*}(b^n)} Q_{L\hat{B}^n}(l, b^n) \overset{(73)}{\leq} \sqrt[4]{2Me^{-n\delta^2\mu}}. \tag{74}$$

Thus by Lemma 4, we see that given a random seed $S \sim \mathsf{unif}(\llbracket 1, 2^{n\rho} \rrbracket)$ for all $b^n \in \mathcal{G}_{\mathcal{C}^*}$, we can construct $f_{b^n} : \llbracket 1, 2^{n\rho} \rrbracket \to \llbracket 1, 2^{nR} \rrbracket$ with

$$\| Q_{f_{b^n}(S)} - Q_{L|\tilde{B}^n = b^n} \|_1 \leq \frac{|\mathcal{L}_{C^*}(b^n)|}{2^{n\rho}} + \sqrt[4]{2Me^{-n\delta^2\mu}} \tag{75}$$

$$\overset{(73)}{\leq} 2^{1+n(R-I(A;B)+4\delta \log_2 M-\rho)} + \sqrt[4]{2Me^{-n\delta^2\mu}} \tag{76}$$

$$\overset{(65)}{\leq} \eta \triangleq 2^{1-n\varepsilon} + \sqrt[4]{2Me^{-n\delta^2\mu}}. \tag{77}$$

We can now glue these functions to define

$$\Lambda_{\mathcal{C}^*}(b^n, S) \triangleq \begin{cases} f_{b^n}(S) & b^n \in \mathcal{G}_{\mathcal{C}^*} \\ l^{*k} & b^n \notin \mathcal{G}_{\mathcal{C}^*} \end{cases}, \tag{78}$$

where $l^* \in [\![1, 2^{n\rho}]\!]$. By construction, for the selected code $\mathcal{C}^*$, we now have

$$\sum_{b^n \in \mathcal{G}_{\mathcal{C}^*}} Q_{\tilde{B}^n}(b^n) \parallel Q_{\Lambda_{\mathcal{C}^*}(b^n,S)} - Q_{L|\tilde{B}^n = b^n} \parallel_1 \le \eta,$$

$$\sum_{b^n \in \mathcal{B}^n} Q_{\tilde{B}^n}(b^n) \parallel Q_{\Lambda_{\mathcal{C}^*}(b^n,S)} - Q_{L|\tilde{B}^n = b^n} \parallel_1 \le \eta + 2\eta_0.$$

Since the RHS does not depend on the choice of $\mathcal{C}^*$ in $\mathcal{F}$,

$$\mathbb{E}\left[\parallel Q_{\Lambda_{\mathcal{C}}(\tilde{B}^n,S),\tilde{B}^n} - Q_{L,\tilde{B}^n} \parallel_1 \big| \mathcal{C} \in \mathcal{F}\right] \le \eta + 2\eta_0.$$

Next. using the fact that the variational distance between two p.m.f.s is at most 2, we also have

$$\mathbb{E}\left[\parallel Q_{\Lambda_{\mathcal{C}}(\tilde{B}^n,S),\tilde{B}^n} - Q_{L,\tilde{B}^n} \parallel_1 \big| \mathcal{C} \notin \mathcal{F}\right] \le 2\mathbb{P}[\mathcal{C} \notin \mathcal{F}].$$

Finally, combining the above two equations and using (71) completes the claim. ∎

Lemma 3: Consider the setup of Lemma 2. Let $\mathcal{L}_{\mathcal{C}}(\cdot)$ be as defined in (67). Then, for n large,

$$\mathbb{E}\left[|\mathcal{L}_{\mathcal{C}}(\tilde{B}^n)|\right] \le 2^{1+n(R-I(A;B)-2\delta \log_2(|\mathcal{A}||\mathcal{B}|))}. \tag{79}$$

Proof: Owing to the random codebook construction,

$$\mathbb{E}\left[|\mathcal{L}_{\mathcal{C}}(\tilde{B}^n)|\right] = \mathbb{E}\left[|\mathcal{L}_{\mathcal{C}}(\tilde{B}^n)| \, \big| \, L = 1\right] \tag{80}$$

$$= \sum_l \mathbb{E}\left[\mathbb{1}\{l \in \mathcal{L}_{\mathcal{C}}(\tilde{B}^n)\} \, \big| \, L = 1\right]. \tag{81}$$

Since the codewords are chosen randomly, it follows that $\mathbb{E}\left[\mathbb{1}\{l \in \mathcal{L}_{\mathcal{C}}(\tilde{B}^n)\}|L=1\right]$ is the same for $l > 2$. Hence,

$$\mathbb{E}\left[|\mathcal{L}_{\mathcal{C}}(\tilde{B}^n)|\right] \le 1 + (2^{nR} - 1)\,\mathbb{E}\left[\mathbb{1}\{2 \in \mathcal{L}_{\mathcal{C}}(\tilde{B}^n)\}|L=1\right]. \tag{82}$$

Clearly, $\mathbb{E}\left[\mathbb{1}\{2 \in \mathcal{L}_{\mathcal{C}}(\tilde{B}^n)\}|L=1\right]$ is exactly the probability that realizations $A^n \sim Q_A^{\otimes n}$, $B^n \sim Q_B^{\otimes n}$ selected independent of one another are jointly δ-letter typical. Thus, by [8, Theorem 1.1], it follows that

$$\mathbb{E}\left[\mathbb{1}\{2 \in \mathcal{L}_{\mathcal{C}}(\tilde{B}^n)\}|L=1\right] = \sum_{(a,b^n) \in T_\delta^n[Q_{AB}]} Q_A(a^n) Q_B(b^n) \le 2^{-n(I(A;B)-2\delta \log_2 |\mathcal{A}||\mathcal{B}|)}.$$

Combining the above bound with (82) completes the proof. ∎

Lemma 4: Let Q be a p.m.f. on a finite set $\mathcal{A}$ such that there exists $\mathcal{B} \subseteq \mathcal{A}$ with $|\mathcal{B}| = M$ and $\sum_{b \in \mathcal{B}} Q(b) \ge 1 - \varepsilon$ for $0 < \varepsilon < 1$. Now, suppose that $L \sim \text{unif}([\![1, \ell]\!])$. Then, there exists $f : [\![1, \ell]\!] \to \mathcal{A}$ such that $Q_{f(L)}$, the p.m.f. of $f(L)$, satisfies $\parallel Q_{f(L)} - Q \parallel_1 \le \varepsilon + \frac{M}{\ell}$.

Proof: Let $b_1 \preceq b_2 \preceq \cdots \preceq b_M$ be an ordering of $\mathcal{B}$. Let $p_0 = 0$, and for $1 \le i \le M$, let $p_i \triangleq \sum_{j=1}^i Q(b_j)$ denote the cumulative mass function. Now, let $N_i \triangleq \lfloor p_i \ell \rfloor$, $i = 0, \ldots, M$, and let $f : [\![1, N_M]\!] \to \mathcal{B}$ be defined by the pre-images via $f^{-1}(b_i) = \{N_{i-1}+1, \ldots, N_i\}$, $i = 1, \ldots, M$. Fig. 3 provides an illustration of these operations. Now, by construction, we have

$$0 \le p_i - \mathbb{P}\left[f(L) \in \{b_1, \ldots, b_i\}\right] \le \ell^{-1}, \quad i = 1, \ldots, M. \tag{83}$$

Consequently, we also have for any $i = 1, \ldots, M$,

$$-\ell^{-1} \le p_i - p_{i-1} - Q_{f(L)}(b_i) = Q(b_i) - Q_{f(L)}(b_i) \le \ell^{-1}. \tag{84}$$
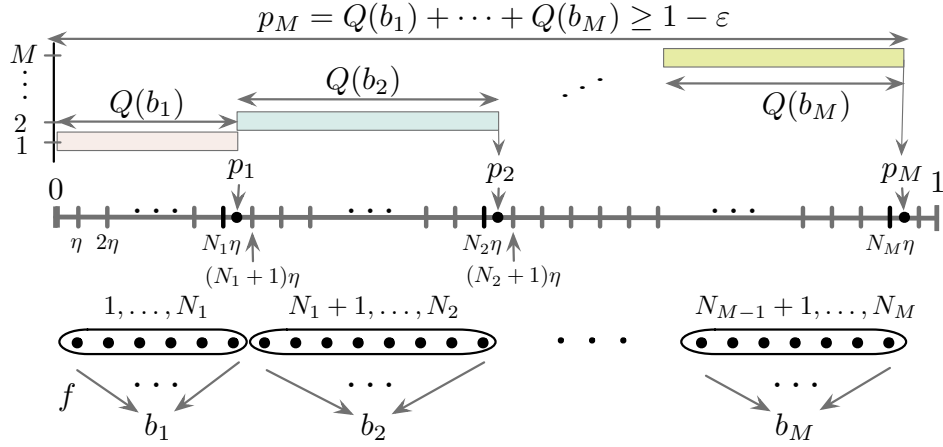
Fig. 3. An illustration of approximating a p.m.f. using a function of a uniform RV.

Hence, we see that

$$\sum_{a \in \mathcal{A}} |Q(a) - Q_{f(L)}(a)| = \sum_{i=1}^{M} |Q(b_i) - Q_{f(L)}(b_i)| + \mathbb{P}[A \notin \mathcal{B}] \overset{(84)}{\leq} \frac{M}{\ell} + \varepsilon. \tag{85}$$

∎

Lemma 5: Given pmf p_{AB} and rates $R_A, R_B \in [0, \infty)$ such that $R_B > I(C; B)$ and $R_A + R_B > I(C; A, B)$, let us construct a random codebook $\{B^n(1), \ldots, B^n(2^{nR_B})\}$ with codewords chosen i.i.d. using $p_B^{\otimes n}$. For each $i \in [\![1, 2^{nR_B}]\!]$, generate a random codebook $\{A^n(i, 1), \ldots, A^n(i, 2^{nR_A})\}$ with codewords chosen i.i.d. using $Q_{A|B}^{\otimes n}(\cdot|B^n(i))$. Let $(I, J) \sim \mathsf{unif}([\![1, 2^{nR_B}]\!] \times [\![1, 2^{nR_A}]\!])$, and let $\hat{C}^n$ be the output when $(A^n(I, J), B^n(I))$ is sent over the DMC $Q_{C|AB}$. Then,

$$\mathbb{E}[D_{KL}(Q_{\hat{C}^n} \| \mathsf{Q}_C^{\otimes n})] \overset{n \to \infty}{\longrightarrow} 0, \tag{86}$$

where the expectation is over all the codebook realizations.

Proof: The proof follows from the achievability scheme and (10)-(15) in [9] by setting $\mathsf{h} = 2$, $A_0 = B$, $A_1 = A$, $X_1 = C$, and $B_1 = B_2 = X_2 = \text{const}$. ∎

Lemma 6 (Theorem 2.2.2 [6]): Let X^n be i.i.d. according to p_X. Then, for each $R < H(X)$, there exists a sequence of mappings $\{\phi_{n,R} : \mathcal{X}^n \to [\![1, 2^{nR}]\!]\}_{n \in \mathbb{N}}$ such that

$$\lim_{n \to \infty} \mathbb{V}(\phi_{n,R}(X^n), \mathsf{unif}([\![1, 2^{nR}]\!])) = 0. \tag{87}$$

REFERENCES

[1] T. S. Han, "Folklore in source coding: Information-spectrum approach," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 747–753, February 2005.
[2] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4619–4637, October 2008.
[3] R. Chou and M. Bloch, "Data compression with nearly uniform output," *2013 IEEE International Symposium on Information Theory*, pp. 1979–1983, 2013.
[4] B. N. Vellambi, M. Bloch, R. Chou, and J. Kliewer, "Lossless and lossy source compression with near-uniform output: Is common randomness always required?" *2015 IEEE International Symposium on Information Theory*, pp. 2171–2175, 2015.
[5] I. Kontoyinannis, "Pointwise redundancy in lossy data compression and universal lossy data compression," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 136–152, January 2000.
[6] T. S. Han, *Information-Spectrum Methods in Information Theory*, 1st ed. Springer, 2003.
[7] P. Cuff and E. Song, "The likelihood encoder for source coding," in *2013 IEEE Information Theory Workshop*, Sept 2013, pp. 1–2.
[8] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.
[9] B. N. Vellambi, J. Kliewer, and M. Bloch, "Strong coordination over multihop line networks," *2015 IEEE Information Theory Workshop*, pp. 192–196, 2015.