Finite Block-Length Gains in Distributed Source Coding

Farhad Shirani
EECS Department
University of Michigan
Ann Arbor,USA
Email: fshirani@umich.edu

S. Sandeep Pradhan
EECS Department
University of Michigan
Ann Arbor,USA
Email: pradhanv@umich.edu

***Abstract*—A new coding scheme for the distributed source coding problem for general discrete memoryless sources is presented. The scheme involves a two-layered coding strategy, the first layer code is of constant finite block-length whereas the second layer contains codes of block-length approaching infinity. It is argued that small block-length codes preserve correlations between sources more efficiently, while suffering rate-loss in a point-to-point compression perspective. Consequently, there is a sweet-spot for the length of the code. An achievable rate-distortion region is characterized using single-letter distributions. It is shown that this region strictly contains previous known achievable rate-distortion regions for the distributed source coding problem.**

I. Introduction

Due to the inherent difficulty in characterization of fundamental limits of reliable communication, most of the works in information theory have considered the analysis of large block-length codes, without any constraints on memory and computational complexity. This allows operating in the realm of the laws of large numbers and thus significantly simplifies the analysis. However, coding theory deals with codes with short block-lengths with computationally efficient encoding and decoding algorithms. While there is interest in application of such small-length communications strategies, it is largely assumed that performance improves when the block-length increases. This turns out to be not true in certain coding schemes used in distributed source coding as discussed below.

The two user distributed source-coding problem is depicted in figure 1. Two correlated discrete-memoryless-sources (DMS) are fed to the encoders. Each encoder compresses its respective source and transmits the compressed version to the decoder. The decoder reconstructs each of the sources with respect to a distortion criterion.There has been a substantial effort to derive the optimal RD region for this set-up, however the region has been characterized only in special cases [2], [3].

The best known inner bound to the optimal rate-distortion (RD) region for distributed coding of two correlated sources is the Berger-Tung (BT) bound [4]. The BT bound is based on a coding strategy called quantize and bin. In this strategy the two sources are quantized using two independent, infinite length random vector quantizers and the quantizers are binned to

This work was supported by NSF grant CCF-1111061.

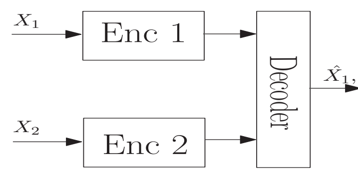

Fig. 1. The two-user distributed source coding problem

reduce the transmission rate. The independent quantization approach leads to the so-called long Markov chain. The Markov chain implies that conditioned on the sources, the single-letter distribution of the quantized versions of the sources decompose into the product of conditional marginal distributions. In [5] it was pointed out that in the presence of common components, further correlation can be induced between the quantized versions, in other words the Markov chain can be relaxed using the common component (CC). Based on this observation the authors in [5] propose a coding scheme which outperforms the BT strategy in the presence of common components, but reduces to the latter in their absence. The CC achievable RD region shrinks discontinuously in source probability distribution as common components are replaced with highly correlated components. From the continuity of the optimal RD region, it was proved in [5] that the CC scheme is also sub-optimal since it is discontinuous. Hence the optimal RD region strictly contains the CC region, however it was not clear how to achieve points outside of the CC rate-distortion region. Toward improving the CC bound, we used two identical random quantizers instead of two independent ones. We found [1] that when applied to the binary-one-help-one (BOHO) example of [5], as the block-length is increased, the performance improves initially, reaches a plateau, and then decreases after that. So the best performance is achieved when the length of the code is at some finite value. Based on this observation, a coding strategy was proposed for the BOHO example which improves the CC achievable RD region.

The main difficulty in analyzing finite block-length coding strategies is that in the absence of simplifying theorems such as laws of large numbers, the resulting characterizations of achievable inner bounds are in terms of multi-letter prob-

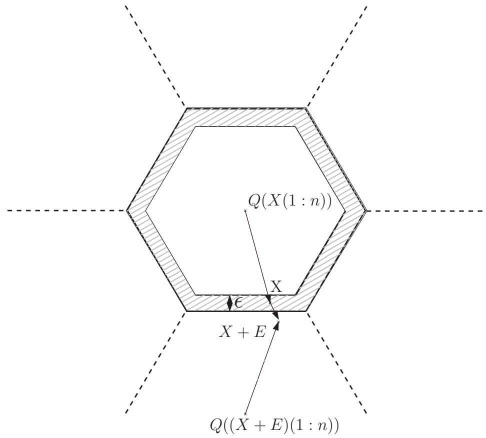

Fig. 2. Quantization noises become independent at large block-length

ability distributions. This makes the computation of such inner bounds very complex. In this paper we first present a coding scheme for the distributed source coding problem in the general discrete, memoryless setting. The scheme utilizes both small-length codes and codes with block-length approaching infinity. A multi-letter characterization of the rate-distortion region achievable using this scheme is given, and it is shown that there is an approximating single-letter characterization for the inner bound. We prove that the resulting inner bound outperforms the previous known coding schemes for this communication setting.

The rest of the paper is organized as follows: In section II, we explain the intuitive reason behind our coding scheme. Section III includes the new coding strategy and the resulting RD region. Section IV concludes the paper.

II. Binary-One-Help One Example

In [1] quantization of two highly correlated BSS's using the same randomly generated code was considered. More specifically, let X be a BSS, and E be a $Be(\epsilon)$ random variable. Define the sources as $X_1 = X$ and $X_2 = X + E$. We took a randomly chosen n-length quantizer Q and quantized X_1^n and X_2^n. As the length of the quantizer approaches infinity, the quantization noises become independent; this happens irrespective of the value of ϵ. The intuitive reason is that as the quantizer length increases, most of the vectors are concentrated on the edges of the Voronoi regions. Hence as the length goes to infinity, with high probability the two source vectors fall into two different quantization regions. Since codewords are pairwise independent, the quantization noises become independent. This is illustrated in figure 2. Independent quantization noises prevent encoders from refining each other's quantizations. When the sources are exactly equal (i.e. common components), the quantizations would also be equal. We argue that the discontinuity in the CC scheme is caused by the discontinuity in the correlation between quantization noises. This implies application of small-length quantizers is beneficial since they preserve correlation between the sources, however there is also a rate-loss associated with using small-length quantizers. The evidence suggests that the existence of a trade-off between avoiding rate-loss and preserving correlation. It was shown that there is a sweet-spot for the length of the quantizer where the trade-off is optimized. Based on these observations a new coding scheme for the special case of BOHO problem was presented. The scheme uses two layers of codes, the first layer consists of a codebook with finite length (i.e. the codeword length is not taken to infinity), and the second codebook has large block-length. In this coding strategy the encoders apply the finite-length quantizer in first layer to quantize the highly correlated components of the sources, and these quantizations are transmitted to the decoder. Since the compressed sequences are highly correlated, the encoders are able to cooperate in transmitting them. In the next step, the CC scheme is applied while considering the previous quantizations as side information completely known at the decoder and partially known at each encoder.

III. The New Coding Strategy

Let X_1 and X_2 be two correlated DMS's. Assume there exist functions $f_1 : \mathcal{X}_1 \to \mathcal{Z}$ and $f_2 : \mathcal{X}_2 \to \mathcal{Z}$, such that $P(f_1(X_1) \neq f_2(X_2)) \leq \epsilon', \epsilon' \in (0, 0.5)$. Here $\mathcal{X}_i$ and $\mathcal{Z}$ are the underlying alphabets for X_i and Z. Let $\epsilon = 1 - (1 - \epsilon')^n$. Also define $S_i = f_i(X_i), i \in \{1, 2\}$. The next theorem presents the main result of this paper:

Theorem 1: The following RD vectors are achievable:

$$
\begin{aligned}
R_1 &\geq I(X_1; U_1|U_2W) + E_{n,\epsilon} + 2|\mathcal{X}_1||\mathcal{U}_1| \log(\frac{p_1}{p_1 - \epsilon}) \\
R_2 &\geq I(X_2; U_2|U_1W) + 2E_{n,\epsilon} + 2|\mathcal{X}_2||\mathcal{U}_2| \log(\frac{p_2}{p_2 - \epsilon}) \\
R_1 + R_2 &\geq I(X_1X_2; U_1U_2W) + 3E_{n,\epsilon} \\
&\qquad + 2|\mathcal{X}_1||\mathcal{U}_1| \log(\frac{p_1}{p_1 - \epsilon}) + \theta_n \\
D_i &\geq E\{d_i\big(h_i(U_1, U_2, W), X_i\big)\}
\end{aligned}
$$

For every distribution P_{X_1,X_2,W,U_1,U_2} satisfying the following constraints:

$$
\begin{aligned}
&1) U_1 \leftrightarrow (X_1, W) \leftrightarrow (X_2, W) \leftrightarrow U_2 \\
&2) W \leftrightarrow S_1 \leftrightarrow (X_1, X_2) \\
&3) p_i > \epsilon, i = 1, 2
\end{aligned}
$$

Also p_i and $E_{n,\epsilon}$ are defined as follows:

$$
\begin{aligned}
p_1 &= \min_{x_1,w,u_1,u_2} (\{P_{U_1|X_1,W,U_2}\}, \{P_{U_1|W,U_2}\}) \\
p_2 &= \min_{x_2,w,u_1,u_2} (\{P_{U_2|X_2,W,U_1}\}, \{P_{U_2|W,U_1}\}) \\
E_{n,\epsilon} &= \frac{h(\epsilon)}{n} + \epsilon \log |\mathcal{W}|
\end{aligned}
$$

In the above formulas, θ_n is a sequence approaching 0 which can be bounded for a given $P_{S,W}$. Also $\mathcal{U}_i$ and $\mathcal{W}$ are the alphabets for U_i and W.

Remark 1: The above RD region reduces to the CC region when $\epsilon = 0$ (i.e when $S_1 = S_2$). Also if S_1 and S_2 are taken to be trivial, the bound reduces to the BT region.

Remark 2: As n becomes larger, ϵ increases, which in turn causes $E_{n,\epsilon}$ to increase. On the other hand, θ_n is a decreasing function of n. This illustrates the trade-off between rate-loss due to application of small block-length codes θ_n, and the gains from preservation of correlation between the sources $E_{n,\epsilon}$.

Remark 3: Note that finding the achievable rate-distortion region involves sweeping over all possible choices of f_1 and f_2. However $\log(\frac{p_1}{p_1-\epsilon})$ increases as f_1 and f_2 become less correlated (i.e. as ϵ increases). This suggests choosing highly correlated functions gives larger achievable regions.

Remark 4: The above inner bound is not symmetric with respect to the two encoders, one can symmetrize the region by swapping the indices for encoders 1 and 2 in the theorem and taking the union of the two resulting regions.

To prove theorem 1, an achievable RD region using finite-length coding schemes is derived, then it is shown that the region contains the inner bound in the theorem.

Let W_1 be a random variable with alphabet $\mathcal{W}_1$. Take an arbitrary probability distribution P_{S_1,W_1}. Also let I be a random variable uniformly distributed on $\{1,2,3,...,n\}$. Consider Q_n, an n-length quantizer which quantizes S_1^n to W_1^n such that:

$$P_{S_1(I),W_1(I)} = \frac{1}{n}\sum_{i\in[1:n]} P_{S_1(i),W_1(i)} = P_{S_1,W_1}.$$

By random coding arguments Q_n exists with rate $R_n = I_{P_{S_1,W_1}}(S_1;W_1) + \theta_n$, where θ_n can be bounded given the distribution P_{S_1,W_1} and approaches 0 as n goes to infinity [6]. Let $W_2^n = Q_n(S_2^n)$. W_2^n can be perceived as the second encoder's "estimate" of W_1^n. We have:

$$P_{X_1^n,X_2^n,W_1^n,W_2^n} = \sum_{s_1^n,s_2^n} P_{X_1^n,X_2^n|S_1^n,S_2^n} P_{S_1^n,S_2^n,W_1^n,W_2^n}$$

Define $P_{X_1,X_2,W_1,W_2} = P_{X_1(I),X_2(I),W_1(I),W_2(I)}$. Also define $\mathcal{P}$ as the set of all probability distributions on X_1,X_2,W_1,W_2,U_1,U_2 such that P_{X_1,X_2,W_1,W_2} is produced by the above process and U_1 and U_2 satisfy $U_1 \leftrightarrow (X_1,W_1) \leftrightarrow (X_2,W_2) \leftrightarrow U_2$. The ensuing theorem states the n-letter achievable bound for the new coding strategy.

Theorem 2: RD vectors satisfying the following bounds are achievable:

$$\begin{aligned}
&R_1 \geq I(X_1;U_1|U_2W_1W_2) + E_{n,\epsilon}\\
&R_2 \geq I(X_2;U_2|U_1W_1W_2) + E_{n,\epsilon}\\
&R_1 + R_2 \geq I(W_1;S_1) + I(X_1;U_1|W_1,W_2)+\\
&\qquad I(X_2;U_2|W_1,W_2) - I(U_1;U_2|W_1W_2) + \theta_n + E_{n,\epsilon}\\
&D_i \geq E\{d_i\big(h_i(U_1,U_2,W_1,W_2),X_i\big)\}
\end{aligned}$$

For every probability distribution $P_{X_1,X_2,W_1,W_2,U_1,U_2}$ chosen form $\mathcal{P}$. Here $h_i : \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathcal{X}_i$ are the reconstruction functions at the decoder.

Remark 5: Q_n completely determines P_{X_1,X_2,W_1,W_2}, also from the Markov chain $P_{U_1|X_1,W_1}$, $P_{U_2|X_2,W_2}$ and Q_n fix the induced joint probability distribution $P_{X_1,X_2,W_1,W_2,U_1,U_2}$. Hence determining the RD region given in theorem 2 involves taking the union of RD vectors satisfying the above bounds for some given f_i, h_i, Q_n, $P_{U_1|X_1,W_1}$ and $P_{U_2|X_2,W_2}$.

Proof: First we present a summary of the proof. Fix f_i, h_i, Q_n, $P_{U_1|X_1,W}$ and $P_{U_2|X_2,W_2}$. Using Q_n the encoders quantize S_i^n to W_i^n. From random coding arguments, if multiple realizations of W_1^n's are available at the decoder, encoder 2 can transmit the corresponding sequence of W_2^n's using rate at most $\frac{1}{n}H(W_2^n|W_1^n)$. So the encoders transmit the sequences of W_i's with sum-rate less than $R_n + \frac{1}{n}max\{H(W_2^n|W_1^n), H(W_1^n|W_2^n)\}$. Since S_i's are highly correlated, the vectors S_i^n are almost always equal. Consequently the quantizations W_i^n are almost always equal, and using this the term $\frac{1}{n}max\{H(W_2^n|W_1^n), H(W_1^n|W_2^n)\}$ can be bounded. In the next step (W_i,X_i) is transformed into a DMS by applying the interleaving method explained in [1]. The rest of the problem can be viewed as distributed source coding with sources (W_i,X_i) and side-information (W_1,W_2) available at the decoder. A more rigorous proof is presented next.

Codebook Generation: Three codebooks are used for the quantization. C^n is the underlying codebook for Q_n. The other two codebooks C_i^m, are constructed by choosing each of their elements from $\mathcal{U}_i$ based on distribution P_{U_i}. C_i^m have rates $I(X_i,W_i;U_i)+\lambda_m$ where $\lambda_m \rightarrow 0$. Let $Q_{i,m}$ be the quantizers associated with these codebooks. Each of the codebooks C_i^m are randomly binned at rate $I(X_i;U_i|W_1,W_2) - r_i$, where $r_1 + r_2 = I(U_1;U_2|W_1,W_2)$. Given $t \in [0,1]$, randomly and uniformly bin the space of all vectors $W^{tnm} \in \mathcal{W}^{tnm}$ with rate $E_{n,\epsilon}$, and define B_1 as the binning function. Also bin the space of all vectors $W^{(1-t)nm}$ using the same rate, let B_2 be the binning function. Finally choose m permutations $\pi_j, j \in [1:m]$ randomly and uniformly from the set of all n-length permutations S_n.

Encoding: Communication is carried out over blocks of length mn. Denote the source sequence in one block as the matrix $X_i(1:m,1:n)$. The ith encoder calculates $W_i(j,1:n) = Q_n(S_i(j,1:n))$ for all j. The encoder wishes to utilize the codebooks C_i^m for quantizing (X_i,W_i), however the source (X_i,W_i) is not a DMS since W_i is produced by a finite-length quantizer. To overcome this difficulty we use the method explained in [1]. Let $\tilde{X}_i(j,1:n) = \pi_j(X(j,1:n))$, define $\tilde{W}_i$ in the same manner. As shown in [1], $(\tilde{X}_i,\tilde{W}_i)(1:m,l)$ would behave like a DMS with probability distribution P_{X_i,W_i}. Each encoder calculates $\tilde{U}_i(1:m,l) = Q_{i,m}((\tilde{X}_i,\tilde{W}_i)(1:m,l))$. For rows $(1:tm)$, the first encoder transmits $W_1(j,1:n)$ while the second encoder sends the bin index $B_1(W_2(1:tm,1:n))$. For the rest of the rows encoder 1 sends the bin index $B_2(W_1(tm+1:m,1:n))$ while encoder 2 sends $W_2(j,1:n)$. In other words the encoders time-share between two strategies. In the first strategy encoder 1 transmits W_1 while encoder 2 only sends the bin number for the sequence of W_2, in the second strategy the encoders reverse roles. For every column l, the i_{th} encoder also sends the bin index of $\tilde{U}_i(1:m,l)$ in

C_i^m. The resulting rates are:

$$R_1 = tR_n + (1-t)E_{n,\epsilon} + I(X_1;U_1|W_1W_2) - r_1$$
$$R_2 = (1-t)R_n + tE_{n,\epsilon} + I(X_2;U_2|W_1W_2) - r_2.$$

If we prove that the above rates are achievable, the proof is complete, since they include both corner points of the region in theorem 2.

Decoding: The decoder first decodes $W_i(1:m,1:n)$. For elements $(1:tm,1:n)$, W_1^{tmn} are available, while only the bin index of W_2^{tmn} is available. Since m is going to infinity, by random coding arguments W_2^{tmn} is losslessly recovered as long the binning rate is more than $\frac{1}{n}H(W_2^n|W_1^n)$. By lemma 1 in the appendix we have $\frac{1}{n}H(W_2^n|W_1^n) \leq E_{n,\epsilon}$. By the same argument (W_1,W_2) are recovered losslessly for the rest of the rows.

The bin size for each vector $\tilde{U}_1(1:m,l)$ is:

$$\begin{aligned} &I(X_1,W_1;U_1) - I(X_1;U_1|W_1,W_2) + r_1 \\ &= I(X_1,W_1,W_2;U_1) - I(X_1;U_1|W_1,W_2) + r_1 \\ &= I(W_1,W_2;U_1) + r_1 \end{aligned}$$

The long Markov chain is used in the second equation. By the same calculations the bin size for $\tilde{U}_2$ is $I(W_1,W_2;U_2)+r_2$. Using typicality arguments one can show that for these bin sizes, there is only one pair $(U_1,U_2)(1:m,l)$, jointly typical with $(W_1,W_2)(1:m,l)$. Due to space limitations we only present a summary of the proof. The decoder first creates two ambiguity sets $\mathcal{L}_i$ from the sequences of $U_i(1:m,l)$'s in the corresponding bins. Each of these sets contains all sequences $U_i(1:m,l)$ in the bin, which are typical with $(W_1,W_2)(1:m,l)$. There is roughly one such sequence in each $2^{mI(W_1,W_2;U_i)}$ vectors. So the size of $\mathcal{L}_i$ is close to 2^{mr_i}. The decoder finds a pair of vectors in the two ambiguity sets which are typical with each other. Since all these vectors are typical with W_1 and W_2, as long as $r_1+r_2 \leq I(U_1;U_2|W_1,W_2)$ there is no more than one pair $(U_1,U_2)(1:m,l)$ typical with respect to $P_{U_1,U_2|W_1,W_2}$. This completes the proof. ∎

From remark 3, calculation of the RD region in the theorem requires taking union over all possible n-length quantizers. This renders the characterization practically incomputable. The next proof shows that the RD region in theorem 1 is contained in the one in theorem 2, so it is achievable using the finite-length coding scheme.

Proof: First we eliminate W_2 from the inequalities. Restricting the reconstruction functions to only use W_1,U_1,U_2 results in an achievable inner bound. In the next step, W_2 is removed from the mutual information terms:

$$\begin{aligned} &I(X_1;U_1|U_2,W_1,W_2) \\ &= H(U_1|U_2,W_1,W_2) - H(U_1|X_1,W_1,U_2) \\ &\leq I(U_1;X_1|W_1,U_2) \end{aligned}$$

Also:

$$\begin{aligned} &I(X_2;U_2|U_1,W_1,W_2) \\ &\leq I(X_2;W_2,U_2|W_1,U_1) \\ &\leq I(X_2;U_2|W_1,U_1) + H(W_2|W_1) \\ &\leq I(X_2;U_2|W_1,U_1) + E_{n,\epsilon} \end{aligned}$$

W_2 in the terms $I(X_1;U_1|W_1,W_2)$ and $I(X_2;U_2|W_1,W_2)$ in the sum-rate bound can be removed using the same method. For $I(U_1;U_2|W_1,W_2)$ an upper-bound is necessary:

$$\begin{aligned} &I(U_1;U_2|W_1,W_2) \\ &\geq I(U_1;U_2|W_1) - H(W_2|W_1) \\ &\geq I(U_1;U_2|W_1) - E_{n,\epsilon} \end{aligned}$$

It is straightforward to show $I(W_1;S_1) = I(W_1;X_1)$. So far we have the following inner bound:

$$\begin{aligned} &R_1 \geq I(X_1;U_1|W_1,U_2) + E_{n,\epsilon} \\ &R_2 \geq I(X_2;U_2|W_1,U_1) + 2E_{n,\epsilon} \\ &R_1+R_2 \geq I(W_1;X_1) + 3E_{n,\epsilon} + I(X_1;U_1|W_1) \\ &\qquad + I(X_2;U_2|W_1) - I(U_1;U_2|W_1) + \theta_n \\ &D_i \geq E\{d_i\big(h_i(U_1,U_2,W_1),X_i\big)\} \end{aligned}$$

Q_n is still playing a role in the calculation of the RD region by determining P_{X_1,X_2,W_1,W_2}, which in turn affects $P_{U_2|X_1,X_2,W_1}$. Lemma 2 in the appendix provides the means to eliminate this dependency on Q_n. The lemma shows that for every probability distribution in theorem 1, there is a probability distribution in theorem 2, for which the above bounds are well-approximated when calculated using one of the distributions instead of the other. Hence we can provide an inner bound for the above rates by considering the distributions from the first theorem and bounding the estimation error using lemma 2. Applying the estimation bounds gives the region presented in theorem 1. ∎

Finally, we show that the RD region in theorem 1 strictly contains the CC rate-distortion region.

Theorem 3: For the BOHO problem in [1], the RD region in theorem 2 achieves points outside of the CC rate-distortion region.

Proof: Take $U_1 = \phi$, $U_2 = Z + X + W + N_{\delta_0}$ and $W = X + N_\epsilon + N_\delta$, where N_δ is $Be(\delta)$,N_{δ_0} is $Be(\delta_0)$, and the quantization noises are independent of the sources and each other. The reconstruction function is $U_2 + W = X + Z + N_{\delta_0}$. Consider the corner point where encoder 1 is transmitting W by itself and encoder 2 is binning its correlated quantization at rate $E_{n,\epsilon}$. The resulting RD vector approaches $(R_1,R_2,D) = (1-h_b(\delta), h_b(p*\delta) - h_b(\delta_0), \delta_0)$ as $\epsilon' \to 0$ and $n \to \infty$. In [5] it was shown that these RD vectors are not achievable by the CC scheme when $\epsilon' \neq 0$. By the same argument as in [1], it can be proved that there exist n and ϵ' for which the resulting rates are not achievable by the CC scheme. ∎

There can be highly correlated components between the sources given W_1 and W_2. In this case, there must be several finite-length codebooks super-imposed on each other, one for each of the highly correlated components. The inner bound presented here can be extended to include these new layers.

IV. Conclusion

A new coding scheme for the distributed source coding problem was presented. The scheme used a combination of finite length and large block-length codes. The application of finite length quantizers was justified by showing that they preserve correlation between sources more efficiently. The achievable region for the coding scheme was derived. A single-letter characterization of an inner bound to the achievable region was presented. It was shown the inner bound strictly contains previous known achievable RD regions.

Appendix

Lemma 1: Let S_1 and S_2 be two DMS's such that $P(S_1 \neq S_2) \leq \epsilon'$ for some $\epsilon' > 0$. Also Let $W_i^n = f_i(S_i^n)$ be n-letter functions of S_i^n to alphabet $\mathcal{W}^n$. Let $\epsilon = 1-(1-\epsilon')^n$. Then the following are true:

1) $P(W_1^n \neq W_2^n) \leq \epsilon$
2) $\frac{1}{n}H(W_2^n|W_1^n) \leq \frac{h(\epsilon)}{n} + \epsilon \log |\mathcal{W}_2|$

Proof: 1)

$$P(W_1^n = W_2^n) \geq P(S_1^n = S_2^n) = 1-\epsilon$$

The first inequality is true since if $S_1^n = S_2^n$, then $W_1^n = W_2^n$.

2) Let 1_A be the indicator function of event A. We have:

$$\begin{aligned}
&H(W_2^n|W_1^n) = H(1_{\{W_1^n=W_2^n\}}, W_2^n|W_1^n)\\
&\leq H(1_{\{W_1^n=W_2^n\}}) + H(W_2^n|W_1^n, 1_{\{W_1^n=W_2^n\}})\\
&\leq h_b(\epsilon) + P(W_1^n \neq W_2^n)H(W_2^n|W_1^n, 1_{\{W_1^n=W_2^n\}}=0)\\
&\leq h_b(\epsilon) + P(W_1^n \neq W_2^n)H(W_2^n)\\
&\leq h_b(\epsilon) + \epsilon n \log|\mathcal{W}_2|.
\end{aligned}$$

■

Lemma 2: Consider a probability distribution P_{X_1,X_2,W_1,U_1,U_2} satisfying the Markov chains $U_1 \leftrightarrow (X_1,W_1) \leftrightarrow (X_2,W_1) \leftrightarrow U_2$ and $W_1 \leftrightarrow S_1 \leftrightarrow (X_1,X_2)$, where S_1 and S_2 are as defined previously. let P_{S_1,W_1} be the marginal distribution of (S_1,W_1), take a quantizer Q_n from $\mathcal{Q}_{P_{S,W}}$. Assume ${P'}_{X_1,X_2,W_1,W_2}$ is the probability distribution induced by Q_n. Let ${P'}_{U_1|X_1,W_1} = P_{U_1|X_1,W_1}$ and ${P'}_{U_2|X_2,W_2} = P_{U_2|X_2,W_1}$. Define:

$${P'}_{X_1,X_2,W_1,W_2,U_1,U_2} = {P'}_{X_1,X_2,W_1,W_2}{P'}_{U_1|X_1,W_1}{P'}_{U_2|X_2,W_2}.$$

The following hold:

$$\begin{aligned}
&1) {P'}_{X_1,X_2,W_1,U_1,U_2} \doteq P_{X_1,X_2,W_1,U_1,U_2} \pm \epsilon\\
&2) I_P(X_1;U_1|W_1,U_2) \doteq I_{P'}(X_1;U_1|W_1,U_2)\\
&\qquad\qquad \pm 2|\mathcal{X}_1||\mathcal{U}_1| \log\left(\frac{p_1}{p_1-\epsilon}\right)\\
&3) I_P(X_2;U_2|W_2,U_1) \doteq I_{P'}(X_2;U_2|W_2,U_1)\\
&\qquad\qquad \pm 2|\mathcal{X}_2||\mathcal{U}_2| \log\left(\frac{p_2}{p_2-\epsilon}\right)
\end{aligned}$$

In the above equations, $a \doteq b \pm \epsilon$ means $a \in [b-\epsilon, b+\epsilon]$. Also $I_P(A;B|C)$ denotes the mutual information with respect to P.

Proof: 1)

$$\begin{aligned}
&{P'}_{X_1,X_2,W_1,U_1,U_2}\\
&= \sum_{w_2} {P'}_{X_1,X_2,W_1,W_2,U_1,U_2}\\
&\geq \sum_{w_2} 1_{\{W_1=W_2\}} {P'}_{X_1,X_2,W_1,W_2,U_1,U_2}\\
&= \sum_{w_2} 1_{\{W_1=W_2\}} {P'}_{X_1,X_2,W_1,W_2} P_{U_1|X_1,W_1} P_{U_2|X_2,W_1}\\
&= P_{U_1|X_1,W_1} P_{U_2|X_2,W_1} \sum_{w_2} 1_{\{W_1=W_2\}} {P'}_{X_1,X_2,W_1,W_2}\\
&\overset{a}{\geq} P_{U_1|X_1,W_1} P_{U_2|X_2,W_1} (P_{X_1,X_2,W_1} - \epsilon)\\
&\geq P_{X_1,X_2,W_1,U_1,U_2} - \epsilon
\end{aligned}$$

Where part (a) results from the following:

$$\begin{aligned}
P_{X_1,X_2,W_1} &= {P'}_{X_1,X_2,W_1}\\
&= \sum_{w_2} {P'}_{X_1,X_2,W_1,W_2}\\
&\leq \sum_{w_2} 1_{\{W_1=W_2\}} {P'}_{X_1,X_2,W_1,W_2} + P(1_{W_1 \neq W_2}).
\end{aligned}$$

Now we prove the other side of the inequality:

$$\begin{aligned}
&{P'}_{X_1,X_2,W_1,U_1,U_2}\\
&= \sum_{w_2} {P'}_{X_1,X_2,W_1,W_2,U_1,U_2}\\
&\leq \sum_{w_2} 1_{W_1=W_2} {P'}_{X_1,X_2,W_1,W_2,U_1,U_2} + \epsilon\\
&\leq P_{X_1,X_2,W_1} P_{U_1|X_1,W_1} P_{U_2|X_1,W_1} + \epsilon\\
&= P_{X_1,X_2,W_1,U_1,U_2} + \epsilon.
\end{aligned}$$

2) and 3) follow from 1 in a straightforward manner by expanding the mutual informations, the maximum difference between the terms in the mutual informations is $2\log\frac{p_i}{p_i-\epsilon}$, for the sake of brevity we omit the proofs for these parts. ■

References

[1] F. Shirani, A.G. Sahebi, S.S. Pradhan, "Distributed source coding in absence of common components," in IEEE International Symp. on Information Theory. IEEE, pp.1362,1366, 7-12 July 2013

[2] D. Slepian, J. Wolf, "Noiseless coding of correlated information sources", IEEE Transactions on Information Theory, vol.19, no.4, pp. 471- 480, Jul 1973

[3] T. Berger, R. W. Yeung, "Multiterminal source encoding with one distortion criterion,", IEEE Transactions on Information Theory, vol. 35, no. 2, pp. 228236, 1989.

[4] S.-Y. Tung, "Multiterminal source coding", Ph.D. Thesis, Cornell University, Ithaca, NY, 1974.

[5] A.B. Wagner, B.G. Kelly, Y. Altuğ, "Distributed Rate-Distortion With Common Components", IEEE Transactions on Information Theory, vol.57, no.7, pp.4035-4057, July 2011

[6] V. Kostina, S. Verdu, "Fixed-Length Lossy Compression in the Finite Blocklength Regime", IEEE Transactions on Information Theory, vol.58, no.6, pp.3309,3338, June 2012