Codebook Information Does Not Reduce Output Entropy when Rate Is above Capacity

Xiugang Wu and Liang-Liang Xie

Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, ON, Canada N2L 3G1
Email: x23wu@uwaterloo.ca, llxie@uwaterloo.ca

Abstract—Under the standard random coding framework, the conditional entropy rate of the channel output given the codebook information is investigated. It is shown that there exists an interesting dichotomy: with respect to the unconditional entropy rate of the output, this conditional entropy rate is reduced when the communication rate is below the channel capacity, but remains the same when the communication rate is above the channel capacity. Hence, the channel capacity plays a role of the critical point. Shannon used the random codebook argument to prove that any communication rate below the channel capacity is achievable, i.e., the destination can successfully decode the transmitted codeword based on the codebook used at the source. However, this cannot be done when the communication rate is above the channel capacity. In this case, our evaluation shows that the codebook information cannot be used to reduce the output entropy rate, and instead, the output entropy rate turns to be the same as if a completely random, instead of a specific codebook, was used at the source.

This characterization of the conditional entropy rate sheds some light on the optimal design of the compress-and-forward relay schemes, where a long standing question is whether the source's codebook information can be used for more effective compressions at the relay. In this paper, specifically, we show that if lossless compression is to be performed, then the codebook information cannot be used to reduce the compression rate at the relay.

I. Introduction

Consider a discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ with the capacity $C := \max_{p(x)} I(X;Y)$. Under the random coding framework, a random codebook $\mathbf{C}_n$ with respect to $p(x)$ with the rate R and block length n is defined as

$$\mathbf{C}_n := \left\{ X^n(w) \in \mathcal{X}^n, w = 1, \ldots, 2^{nR} \right\}, \quad (1)$$

where each codeword in $\mathbf{C}_n$ is an i.i.d. random sequence generated according to some input distribution $p(x)$. Shannon [1] used such random codebooks to show that any communication rate less than the channel capacity is achievable. It is interesting to note that with high probability, such randomly generated codebooks will be "good" in the sense that the decoding error probability will approach zero as $n \to \infty$ when $R < I(X;Y)$. It is also well known that in the case when $R > I(X;Y)$, the decoding error probability will not be small, but instead, approach one as $n \to \infty$.

In this paper, we study the conditional entropy rate of the channel output given the codebook information, namely $\lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n)$. Our main result is that this conditional entropy rate is a piecewise linear function of the communication rate, given by

$$\lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n) = \begin{cases} H(Y) & \text{when } R > I(X;Y), \\ R + H(Y|X) & \text{when } R < I(X;Y), \end{cases} \quad (2)$$

where $H(Y)$, $I(X;Y)$ and $H(Y|X)$ are all calculated according to $p(x,y) = p(x)p(y|x)$. In contrast, without the codebook information, the unconditional entropy rate of the output, namely $\lim_{n\to\infty} \frac{1}{n} H(Y^n)$, equals $H(Y)$ for any $R > 0$. Interestingly, here we see a dichotomy: with the codebook information, the conditional entropy rate reduces when $R < I(X;Y)$ but remains the same when $R > I(X;Y)$, where $I(X;Y)$ plays a role of the critical point.

In this paper, we are especially interested in the case when $R > I(X;Y)$, so that the receiver cannot successfully decode the message. This is motivated by investigating the optimal compress-and-forward schemes for the relay channel. For simplicity, consider the relay channel depicted in Fig. 1, where the source's input X is received by the relay Y and the destination Z through a channel $p(y,z|x)$, and the relay can communicate to the destination via an error-free digital link with rate R_0. When the source is transmitting at a rate
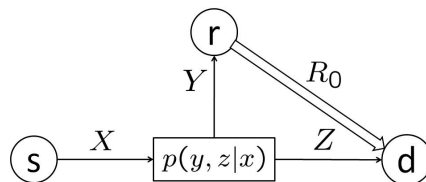

Fig. 1. A relay channel model with a digital link.

higher than $I(X;Y)$, the relay cannot decode the message, but it can help by compressing and forwarding its observations to the destination via the digital link. The optimal design of the compress-and-forward schemes at the relay is arguably one of the most critical problems in the development of network information theory, where ambiguity always arises when decoding cannot be performed successfully. In the classical compress-and-forward scheme [3] and some of its variants [4]-[7], the relay simply treats its observation Y^n as completely randomly generated according to $p(y) = \sum_x p(x)p(y|x)$, and

the compression scheme at the relay is only based on the distribution used for generating the codebook at the source, i.e., $p(x)$, instead of the specific codebook used. While many different codebooks can be generated according to the same distribution, can the relay exploit the structure of the specific codebook used and compress its observations more efficiently?

There have been some discussions on this issue [7]-[11]. Particularly, in [9]-[11], we showed that for a single user channel, if a randomly generated codebook with the rate above capacity is used, then the output satisfies an asymptotic equipartition property (AEP), irrespective of the specific codebook used. Essentially, this insight demonstrates that although each different codebook may have its own structure, this structure will disappear from the output's point of view as long as the rate is above capacity. And this loss of structure suggests that the observations at the relay are somehow independent of the specific codebook used at the source, and only depend on the distribution by which the codebook is generated, when the source's rate is above the capacity of the source-relay link.

As an application of our main result shown in (2), in this paper, we will further explore on this issue. Specifically, we focus on the case where lossless compression is to be performed at the relay and ask the question: what is the minimum required R_0 in Fig. 1 when the relay wants to transfer Y^n faithfully to the destination? We will calculate this minimum required R_0 in two different scenarios: i) the relay uses the knowledge of the source's codebook to do the compression; ii) the relay simply ignores this knowledge by treating Y^n as randomly generated according to $p(y) = \sum_x p(x)p(y|x)$. Surprisingly, it is shown that the minimum required R_0 in both scenarios are the same, no matter whether there exists side information at the destination, which indicates that it is indeed optimal for the relay to compress its observation without using the source's codebook information in the lossless case.

The remainder of the paper is organized as follows. First, the main result of this paper is formally stated in Section II, followed by its proof in Section III. Then, as an application of the result, the optimal design of the compress-and-forward schemes is discussed in Section IV.

II. MAIN RESULT

Theorem 1: Consider a discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$. Under the random coding framework, let $\mathbf{C}_n$ be a random codebook i.i.d. generated according to some $p(x)$ with rate R and block length n. Then, the conditional output entropy rate given the codebook information is

$$\lim_{n \to \infty} \frac{1}{n} H(Y^n | \mathbf{C}_n) = \begin{cases} H(Y) & \text{when } R > I(X;Y), \\ R + H(Y|X) & \text{when } R < I(X;Y), \end{cases}$$
(3)

where $H(Y)$, $I(X;Y)$ and $H(Y|X)$ are all calculated according to $p(x,y) = p(x)p(y|x)$.

III. CONDITIONAL ENTROPY RATE OF OUTPUT GIVEN CODEBOOK INFORMATION

The proof of Theorem 1 is based upon the results obtained in [9]-[11], which essentially discussed the asymptotic equipar-

tition property (AEP) of the output when the rate is above capacity. Thus, before presenting the proof of Theorem 1, we first review some definitions on strong typical sequences, and the useful results in [9]-[11], in III-A and III-B respectively. Then, we characterize the conditional entropy rate of the output given the codebook information in III-C.

A. Strong Typicality

We begin with some standard definitions on strong typicality [12, Ch. 3].

Definition 1: The ϵ-strongly typical set with respect to $p(x)$, denoted by $A_\epsilon^{(n)}(X)$, is the set of sequences $x^n \in \mathcal{X}^n$ satisfying:
1. For all $a \in \mathcal{X}$ with $p(a) > 0$,

$$\left| \frac{1}{n} N(a|x^n) - p(a) \right| < \frac{\epsilon}{|\mathcal{X}|},$$

2. For all $a \in \mathcal{X}$ with $p(a) = 0$, $N(a|x^n) = 0$.
$N(a|x^n)$ is the number of occurrences of a in x^n.

Similarly, we can define the ϵ-strongly typical set with respect to $p(y)$ and denote it by $A_\epsilon^{(n)}(Y)$.

Definition 2: The ϵ-strongly typical set with respect to $p(x,y)$, denoted by $A_\epsilon^{(n)}(X,Y)$, is the set of sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ satisfying:
1. For all $(a,b) \in \mathcal{X} \times \mathcal{Y}$ with $p(a,b) > 0$,

$$\left| \frac{1}{n} N(a,b|x^n, y^n) - p(a,b) \right| < \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|},$$

2. For all $(a,b) \in \mathcal{X} \times \mathcal{Y}$ with $p(a,b) = 0$,

$$N(a,b|x^n, y^n) = 0.$$

$N(a,b|x^n, y^n)$ is the number of occurrences of the pair (a,b) in the pair of sequences (x^n, y^n).

Definition 3: The ϵ-strongly conditionally typical set with the sequence x^n with respect to the conditional distribution $p(y|x)$, denoted by $A_\epsilon^{(n)}(Y|x^n)$, is the set of sequences $y^n \in \mathcal{Y}^n$ satisfying:
1. For all $(a,b) \in \mathcal{X} \times \mathcal{Y}$ with $p(b|a) > 0$,

$$\frac{1}{n} |N(a,b|x^n, y^n) - p(b|a)N(a|x^n)| \leq \epsilon(1 + \frac{1}{|\mathcal{Y}|}), \quad (4)$$

2. For all $(a,b) \in \mathcal{X} \times \mathcal{Y}$ with $p(b|a) = 0$,

$$N(a,b|x^n, y^n) = 0. \quad (5)$$

B. AEP of Output when Rate is above Capacity

We summarize the key definition and results in [9]-[11] as follows.

Definition 4: A codebook

$$\mathcal{C}_n = \left\{ x^n(w) \in \mathcal{X}^n, w = 1, \ldots, 2^{nR} \right\}$$

is said to be ϵ-typical with respect to $p(x)$ if
1) $x^n(w) \in A_\epsilon^{(n)}(X), \forall w \in \{1, \ldots, 2^{nR}\}$,

2) $\sup_{y^n \in A_\epsilon^{(n)}(Y)} \left| \frac{N_\epsilon(y^n|\mathcal{C}_n)}{2^{nR}} - P_\epsilon(y^n) \right| \leq \frac{n^3 R}{2^{nR}}.$

In the above definition,

$$P_\epsilon(y^n) := \Pr(y^n \in A_\epsilon^{(n)}(Y|\tilde{X}^n)|\tilde{X}^n \in A_\epsilon^{(n)}(X)),$$

$$N_\epsilon(y^n|\mathcal{C}_n) := \sum_{w=1}^{2^{nR}} \mathbb{I}(y^n \in A_\epsilon^{(n)}(Y|x^n(w))),$$

where $\tilde{X}^n$ is drawn i.i.d. according to $p(x)$, and $\mathbb{I}(A)$ is the indicator function being 1 if A holds and 0 otherwise.

Proposition 1: Given that an ϵ-typical codebook $\mathcal{C}_n$ is used and the noise is also ϵ-typical, i.e., $Y^n \in A_\epsilon^{(n)}(Y|X^n)$, we have[1]

$$p(y^n|\mathcal{C}_n) \doteq 2^{-nH(Y)}, \forall y^n \in A_\epsilon^{(n)}(Y),$$

when $R > I(X;Y)$.

Proposition 2: Generate the codebook $\mathbf{C}_n$ at random according to $p(x)$ and reserve only the ϵ-strongly typical codewords. Then, for any $\epsilon > 0$,

$$\Pr(\mathbf{C}_n \text{ is } \epsilon\text{-typical}) \to 1 \text{ as } n \to \infty.$$

C. Proof of Theorem 1

To prove Theorem 1, we extend the definition of typical codebooks in [9]-[11] to the generalized typical codebooks.

Definition 5: A codebook

$$\mathcal{C}_n = \left\{ x^n(w) \in \mathcal{X}^n, w = 1, \ldots, 2^{nR} \right\}$$

is said to be ϵ-generalized typical with respect to $p(x)$ if

1) $\mathcal{C}_{n,\epsilon} := \left\{ x^n(w) \in \mathcal{C}_n : x^n(w) \in A_\epsilon^{(n)}(X) \right\}$ is an ϵ-typical codebook,

2) $\left| \dfrac{|\mathcal{C}_{n,\epsilon}|}{2^{nR}} - \Pr(X^n \in A_\epsilon^{(n)}(X)) \right| \le \dfrac{n}{2^{nR/2}}$.

where $\mathcal{C}_{n,\epsilon}$ is a subcode of $\mathcal{C}_n$ containing only the ϵ-strongly typical codewords and X^n is drawn i.i.d. according to $p(x)$.

The following lemma states that a generalized typical codebook still appears with high probability under the standard random codebook construction schema.

Lemma 1: If the codebook $\mathbf{C}_n$ is i.i.d. generated according to $p(x)$, then for any $\epsilon > 0$,

$$\Pr(\mathbf{C}_n \text{ is } \epsilon\text{-generalized typical}) \to 1 \text{ as } n \to \infty.$$

Proof: By Proposition 2, a random codebook satisfies the first condition in Definition 5 with high probability. Using Chebyshev inequality, we can show that a random codebook also satisfies the second condition with high probability and the details are omitted in this version of the paper. ∎

We are ready to prove Theorem 1.

Proof of Theorem 1: We prove Theorem 1 by characterizing $\lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n)$ in two different cases: when $R > I(X;Y)$ and when $R < I(X;Y)$, respectively.

i) When $R > I(X;Y)$

Define an indicator random variable E as

$$E := \mathbb{I}(E_\epsilon),$$

[1]Same as the notation in [12], we say $a_n \doteq b_n$ if $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. "$\dot{\ge}$" and "$\dot{\le}$" have similar interpretations.

where E_ϵ denotes the event $Y^n \in A_\epsilon^{(n)}(Y|X^n)$.

For any generalized typical codebook $\mathcal{C}_n$ and its subcode $\mathcal{C}_{n,\epsilon}$, when $R > I(X;Y)$, we have

$$H(Y^n|\mathcal{C}_n)$$
$$\ge H(Y^n|E, \mathcal{C}_n) \tag{6}$$
$$= \Pr(E=1)H(Y^n|E=1, \mathcal{C}_n) + \Pr(E=0)H(Y^n|E=0, \mathcal{C}_n)$$
$$\ge \Pr(E=1) \cdot H(Y^n|E=1, \mathcal{C}_n)$$
$$= (1-o(1)) \cdot H(Y^n|E=1, \mathcal{C}_n) \tag{7}$$
$$\ge (1-o(1)) \cdot H(Y^n|E_\epsilon, X^n \in \mathcal{C}_{n,\epsilon})\Pr(X^n \in \mathcal{C}_{n,\epsilon}|X^n \in \mathcal{C}_n)$$
$$= (1-o(1)) \cdot H(Y^n|E_\epsilon, X^n \in \mathcal{C}_{n,\epsilon}) \tag{8}$$
$$= (1-o(1)) \cdot \sum_{y^n} p(y^n|E_\epsilon, \mathcal{C}_{n,\epsilon}) \log \frac{1}{p(y^n|E_\epsilon, \mathcal{C}_{n,\epsilon})}$$
$$\ge (1-o(1)) \cdot \sum_{y^n \in A_\epsilon^{(n)}(Y)} p(y^n|E_\epsilon, \mathcal{C}_{n,\epsilon}) \log \frac{1}{p(y^n|E_\epsilon, \mathcal{C}_{n,\epsilon})}$$
$$\ge (1-o(1)) \cdot \sum_{y^n \in A_\epsilon^{(n)}(Y)} p(y^n|E_\epsilon, \mathcal{C}_{n,\epsilon}) \log 2^{n[H(Y)-\epsilon^*]} \tag{9}$$
$$= n[H(Y) - \epsilon^*] \cdot (1-o(1)) \cdot \sum_{y^n \in A_\epsilon^{(n)}(Y)} p(y^n|E_\epsilon, \mathcal{C}_{n,\epsilon})$$
$$= n[H(Y) - \epsilon^*] \cdot (1-o(1)) \cdot \Pr(Y^n \in A_\epsilon^{(n)}(Y)|E_\epsilon, \mathcal{C}_{n,\epsilon})$$
$$= n[H(Y) - \epsilon^*] \cdot (1-o(1))$$
$$\quad \cdot \Pr(Y^n \in A_\epsilon^{(n)}(Y)|E_\epsilon, \mathcal{C}_n, X^n \in A_\epsilon^{(n)}(X))$$

where (6) follows from the fact that conditioning reduces entropy; (7) follows from the fact that $\Pr(E_\epsilon) \to 1$ as $n \to \infty$, for any $\epsilon > 0$; (8) follows from the second condition in the definition of generalized typical codebooks, which basically states that $|\mathcal{C}_{n,\epsilon}|/2^{nR} \to 1$ as $n \to \infty$; (9) follows from Proposition 1, which upper bounds $p(y^n|E_\epsilon, \mathcal{C}_{n,\epsilon})$ by $2^{-n[H(Y)-\epsilon^*]}$ for any $y^n \in A_\epsilon^{(n)}(Y)$, where $\epsilon^* \to 0$ as $n \to \infty$.

Therefore, when $R > I(X;Y)$, we have

$$H(Y^n|\mathbf{C}_n)$$
$$= \sum_{\mathcal{C}_n} p(\mathcal{C}_n) \cdot H(Y^n|\mathbf{C}_n = \mathcal{C}_n)$$
$$\ge \sum_{\mathcal{C}_n \text{ is generalized typical}} p(\mathcal{C}_n) \cdot H(Y^n|\mathbf{C}_n = \mathcal{C}_n)$$
$$\ge n[H(Y) - \epsilon^*] \cdot (1-o(1))$$
$$\quad \cdot \sum_{\mathcal{C}_n \text{ is GT}} p(\mathcal{C}_n) \cdot \Pr(Y^n \in A_\epsilon^{(n)}(Y)|E_\epsilon, \mathcal{C}_n, X^n \in A_\epsilon^{(n)}(X))$$
$$\ge n[H(Y) - \epsilon^*] \cdot (1-o(1))$$
$$\quad \cdot \sum_{\mathcal{C}_n \text{ is GT}} p(\mathcal{C}_n, E_\epsilon, X^n \in A_\epsilon^{(n)}(X))$$
$$\quad\quad \cdot \Pr(Y^n \in A_\epsilon^{(n)}(Y)|E_\epsilon, \mathcal{C}_n, X^n \in A_\epsilon^{(n)}(X))$$
$$\ge n[H(Y) - \epsilon^*] \cdot (1-o(1)) \cdot \Pr(E_\epsilon, X^n \in A_\epsilon^{(n)}(X))$$
$$\quad \cdot \sum_{\mathcal{C}_n \text{ is GT}} p(\mathcal{C}_n|E_\epsilon, X^n \in A_\epsilon^{(n)}(X))$$
$$\quad\quad \cdot \Pr(Y^n \in A_\epsilon^{(n)}(Y)|E_\epsilon, \mathcal{C}_n, X^n \in A_\epsilon^{(n)}(X))$$
$$= n[H(Y) - \epsilon^*] \cdot (1-o(1))$$

$$\cdot \Pr(Y^n \in A_\epsilon^{(n)}(Y), \mathbf{C}_n \text{ is GT}|E_\epsilon, X^n \in A_\epsilon^{(n)}(X))$$
$$= n[H(Y) - \epsilon^*] \cdot (1 - o(1)) \cdot (1 - o(1)) \tag{10}$$
$$= n[H(Y) - \epsilon^*] \cdot (1 - o(1)),$$

where 'GT' is short for 'generalized typical' and (10) follows from Lemma 1, and

$$\lim_{n \to \infty} \frac{1}{n} H(Y^n | \mathbf{C}_n)$$
$$\geq \lim_{n \to \infty} \frac{1}{n} (n[H(Y) - \epsilon^*] \cdot (1 - o(1)))$$
$$= \lim_{n \to \infty} [H(Y) - \epsilon^*] \cdot (1 - o(1))$$
$$= H(Y). \tag{11}$$

Furthermore, it is obvious that

$$\lim_{n \to \infty} \frac{1}{n} H(Y^n | \mathbf{C}_n) \leq \lim_{n \to \infty} \frac{1}{n} H(Y^n) = H(Y). \tag{12}$$

Combining (11) and (12), we have that when $R > I(X;Y)$,

$$\lim_{n \to \infty} \frac{1}{n} H(Y^n | \mathbf{C}_n) = H(Y).$$

ii) When $R < I(X;Y)$

To find $\lim_{n \to \infty} \frac{1}{n} H(Y^n | \mathbf{C}_n)$ when $R < I(X;Y)$, we first introduce two lemmas, whose proofs are omitted in this version of the paper.

Lemma 2: When $R < I(X;Y)$,

$$\frac{1}{n} H(X^n | \mathbf{C}_n, Y^n) \to 0, \text{ as } n \to \infty.$$

Lemma 3:

$$\lim_{n \to \infty} \frac{1}{n} H(X^n | \mathbf{C}_n) = R.$$

Now, expanding $H(X^n, Y^n | \mathbf{C}_n)$ in two different ways, we have

$$H(X^n, Y^n | \mathbf{C}_n) = H(X^n | \mathbf{C}_n) + H(Y^n | X^n, \mathbf{C}_n)$$
$$= H(Y^n | \mathbf{C}_n) + H(X^n | \mathbf{C}_n, Y^n),$$

and thus

$$H(Y^n | \mathbf{C}_n) = H(X^n | \mathbf{C}_n) + H(Y^n | X^n, \mathbf{C}_n) - H(X^n | \mathbf{C}_n, Y^n).$$

Therefore, when $R < I(X;Y)$,

$$\lim_{n \to \infty} \frac{1}{n} H(Y^n | \mathbf{C}_n) = \lim_{n \to \infty} \frac{1}{n} H(X^n | \mathbf{C}_n) + \lim_{n \to \infty} \frac{1}{n} H(Y^n | X^n, \mathbf{C}_n)$$
$$\qquad - \lim_{n \to \infty} \frac{1}{n} H(X^n | \mathbf{C}_n, Y^n)$$
$$= R + \lim_{n \to \infty} \frac{1}{n} H(Y^n | X^n, \mathbf{C}_n) \tag{13}$$
$$= R + \lim_{n \to \infty} \frac{1}{n} H(Y^n | X^n) \tag{14}$$
$$= R + \lim_{n \to \infty} \frac{1}{n} [H(X^n, Y^n) - H(X^n)]$$
$$= R + H(X,Y) - H(X)$$
$$= R + H(Y|X),$$

where (13) follows from Lemma 2 and 3, (14) follows from the fact that $\mathbf{C}_n \to X^n \to Y^n$ forms a Markov Chain. This completes the proof of Theorem 1. ∎

IV. RATE NEEDED TO COMPRESS RELAY'S OBSERVATION

To investigate the optimal compress-and-forward strategy, in this section, we study the rate needed for the relay to losslessly compress its observation. In the classical compress-and-forward strategy [3] and some of its variants [4]-[7], the compression scheme at the relay was only based on the distribution used for generating the codebook at the source, ignoring the specific source's codebook generated. However, since both the relay and destination in the relay channel have the knowledge of the source's codebook, it is natural to ask whether it is beneficial for the relay to compress its observation with source's codebook information used. This question motivates us to compare the rates needed to compress the relay's observation in two different scenarios: when the relay uses the knowledge of the source's codebook and when the relay simply ignores this knowledge.

First, we consider the two compression problems shown in Fig. 2, where the relay's observation Y^n is generated from X^n through the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ and $\mathbf{C}_n$ is the source's codebook information available to both the relay and destination. We show that to perfectly recover Y^n, the minimum required rates in both scenarios, denoted by R_1 and R_2 respectively, are the same when the rate R of $\mathbf{C}_n$ is greater than $I(X;Y)$, i.e., when the relay cannot decode the source's message.
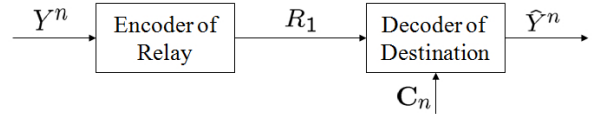
(a) Compression without using source's codebook information
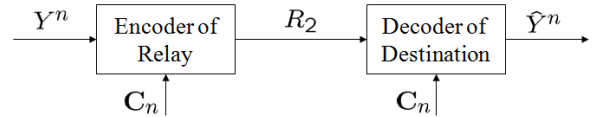
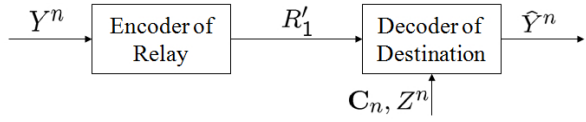(b) Compression using source's codebook information

Fig. 2. Two scenarios where the relay compresses its observation.

Then, we further consider the problems shown in Figure 3, where the destination has access to the side information Z^n, and (Y^n, Z^n) are generated from X^n through the channel $(\mathcal{X}, p(y,z|x), \mathcal{Y} \times \mathcal{Z})$. We show that the minimum required rates to perfectly recover Y^n in both scenarios, denoted by R_1' and R_2' respectively, are still the same, when the rate R of $\mathbf{C}_n$ satisfies
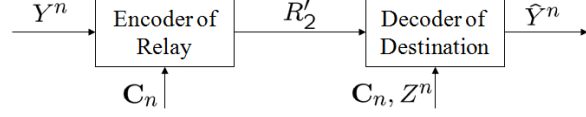
$$I(X;Z) < R < I(X;Y,Z). \tag{15}$$

Note that (15) is a general assumption such that the discussion on the compress-and-forward scheme is meaningful. So our consideration under (15) will not lose any generality.

Formally, we have the following two theorems:

(a) Compression without using source's codebook information

(b) Compression using source's codebook information

Fig. 3. Two scenarios where the relay compresses its observation when there is the side information Z^n

Theorem 2: For the compression problems depicted in Fig. 2, if the rate R of $\mathbf{C}_n$ is larger than $I(X;Y)$, then to perfectly recover Y^n at the destination, the minimum required rates in both scenarios, denoted by R_1 and R_2 respectively, are the same and equal to $H(Y)$.

Theorem 3: For the compression problems depicted in Fig. 3, if the rate R of $\mathbf{C}_n$ satisfies $I(X;Z) < R < I(X;Y,Z)$, then to perfectly recover Y^n at the destination, the minimum required rates in both scenarios, denoted by R_1' and R_2' respectively, are the same and equal to $R - I(X;Z) + H(Y|X,Z)$.

Remark 1: These two theorems imply that, in the lossless case, we cannot compress the relay's observation more efficiently even if the source's codebook information is used, no matter whether there exsits side information at the destination or not.

To prove Theorems 2 and 3, we need the following lemma.

Lemma 4: For the compression problem in Figure 2-(b), we can encode Y^n at rate R_2 and recover it with the probability of error $P_e^{(n)} \to 0$ only if

$$R_2 \geq \lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n). \tag{16}$$

Similarly, for the compression problem in Figure 3-(b), we can encode Y^n at rate R_2' and recover it with the probability of error $P_e^{(n)} \to 0$ only if

$$R_2' \geq \lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n, Z^n). \tag{17}$$

Proof of Lemma 4: The source code for Figure 2-(b) consists of a encoder mapping $f(Y^n, \mathbf{C}_n)$ and a decoder mapping $g(f(Y^n, \mathbf{C}_n), \mathbf{C}_n)$. Let $I = f(Y^n, \mathbf{C}_n)$, then $P_e^{(n)} = \Pr(g(I, \mathbf{C}_n) \neq Y^n)$. By Fano's inequality, for any source code with $P_e^{(n)} \to 0$, we have

$$H(Y^n|I, \mathbf{C}_n) \leq P_e^{(n)} \log |\mathcal{Y}^n| + 1 = P_e^{(n)} n \log |\mathcal{Y}| + 1 = n\epsilon_n, \tag{18}$$

where $\epsilon_n \to 0$ as $n \to \infty$.

Therefore, for any source code with rate R_2 and $P_e^{(n)} \to 0$, we have the following chain of inequalities

$$nR_2 \geq H(I) \tag{19}$$

$$\geq H(I|\mathbf{C}_n)$$
$$= H(Y^n, I|\mathbf{C}_n) - H(Y^n|I, \mathbf{C}_n)$$
$$= H(Y^n|\mathbf{C}_n) + H(I|Y^n, \mathbf{C}_n) - H(Y^n|I, \mathbf{C}_n)$$
$$= H(Y^n|\mathbf{C}_n) - H(Y^n|I, \mathbf{C}_n) \tag{20}$$
$$\geq H(Y^n|\mathbf{C}_n) - n\epsilon_n \tag{21}$$

where (19) follows from the fact that $I \in \{1, 2, \ldots, 2^{nR_2}\}$, (20) follows from the fact that I is a function of Y^n and $\mathbf{C}_n$, and (21) follows from (18). Dividing the inequality $nR_2 \geq H(Y^n|\mathbf{C}_n) - n\epsilon_n$ by n and taking the limit as $n \to \infty$, we establish (16).

Similarly, the source code for Figure 3-(b) consists of a encoder mapping $f'(Y^n, \mathbf{C}_n) := I'$ and a decoder mapping $g'(f'(Y^n, \mathbf{C}_n), \mathbf{C}_n, Z^n)$, and $P_e^{(n)} = \Pr(g'(I', \mathbf{C}_n, Z^n) \neq Y^n)$. By Fano's inequality, for any source code with $P_e^{(n)} \to 0$, we have

$$H(Y^n|I', \mathbf{C}_n, Z^n) \leq P_e^{(n)} \log |\mathcal{Y}^n| + 1 = P_e^{(n)} n \log |\mathcal{Y}| + 1 = n\epsilon_n', \tag{22}$$

where $\epsilon_n' \to 0$ as $n \to \infty$.

Therefore, for any source code with rate R_2' and $P_e^{(n)} \to 0$, we have the following chain of inequalities

$$nR_2' \geq H(I')$$
$$\geq H(I'|\mathbf{C}_n, Z^n)$$
$$= H(Y^n|\mathbf{C}_n, Z^n) + H(I'|Y^n, \mathbf{C}_n, Z^n) - H(Y^n|I', \mathbf{C}_n, Z^n)$$
$$= H(Y^n|\mathbf{C}_n, Z^n) - H(Y^n|I', \mathbf{C}_n, Z^n)$$
$$\geq H(Y^n|\mathbf{C}_n, Z^n) - n\epsilon_n'$$

Again, dividing the inequality $nR_2' \geq H(Y^n|\mathbf{C}_n, Z^n) - n\epsilon_n'$ by n and taking the limit as $n \to \infty$, we establish (17). ∎

Proof of Theorem 2: It is obvious that the minimum required rate R_1 in Fig. 2-(a) and the minimum required rate R_2 in Fig. 2-(b) satisfy that $R_1 \geq R_2$. By treating Y^n as a random outcome generated according to $p(y) = \sum_x p(x)p(y|x)$ in Fig. 2-(a), we know that the rate $H(Y)$ is achievable and hence $R_1 \leq H(Y)$. Besides, by Lemma 4, $R_2 \geq \lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n)$. When the rate R of the codebook $\mathbf{C}_n$ is larger than $I(X;Y)$, we have

$$H(Y) \geq R_1 \geq R_2 \geq \lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n) = H(Y),$$

and thus $R_1 = R_2 = H(Y)$, which proves the theorem. ∎

Proof of Theorem 3:

It has been shown in [4]-[7] that if the destination exploits the structure of the codebook while the relay ignores this structure, then the compression rate $R - I(X;Z) + H(Y|X,Z)$ is achievable, where R is the rate of the source's codebook. Thus, we have $R_1' \leq R - I(X;Z) + H(Y|X,Z)$, where R_1' is the minimum required rate in Fig. 3-(a). Then, noting that $R_1' \geq R_2'$, to show $R_1' = R_2'$, it suffices to show that $R_2' \geq R - I(X;Z) + H(Y|X,Z)$, when $I(X;Z) < R < I(X;Y,Z)$.

By Lemma 4, the minimum required rate R_2' in Fig. 3-(b) should satisfy

$$R_2' \geq \lim_{n\to\infty} \frac{1}{n} H(Y^n|\mathbf{C}_n, Z^n).$$

Applying the result in Theorem 1 for the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ to the channel $(\mathcal{X}, p(y, z|x), \mathcal{Y} \times \mathcal{Z})$, we have $\lim_{n \to \infty} \frac{1}{n} H(Y^n, Z^n | \mathbf{C}_n) = R + H(Y, Z|X)$ when $R < I(X; Y, Z)$. Again, applying Theorem 1 to the channel $(\mathcal{X}, p(z|x), \mathcal{Z})$, we have $\lim_{n \to \infty} \frac{1}{n} H(Z^n | \mathbf{C}_n) = H(Z)$ when $R > I(X; Z)$. Therefore, if $I(X; Z) < R < I(X; Y, Z)$, we have

$$\lim_{n \to \infty} \frac{1}{n} H(Y^n | \mathbf{C}_n, Z^n) = \lim_{n \to \infty} \frac{1}{n} H(Y^n, Z^n | \mathbf{C}_n)$$
$$- \lim_{n \to \infty} \frac{1}{n} H(Z^n | \mathbf{C}_n)$$
$$= R + H(Y, Z|X) - H(Z)$$
$$= R - I(X; Z) + H(Y|X, Z),$$

which concludes the proof of Theorem 3. ∎

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.,* vol.27, pp.379-423, 623-656, 1948.

[2] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 1328–1329 , Mar. 2008.

[3] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, 1979.

[4] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inform. Theory*, vol. IT-52, no. 4, pp. 1545–1561, 2006.

[5] A. El Gamal and Y.-H. Kim, "Lectures notes on network information theory," 2010, available online at http://arxiv.org/abs/1001.3404.

[6] X. Wu and L.-L. Xie, "An optimality-robustness tradeoff in the compress-and-forward relay scheme," in *Proc. IEEE 72nd Vehicular Technology Conference*, Ottawa, Sep. 2010.

[7] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2007.

[8] F. Xue, P. R. Kumar and L.-L. Xie, "The conditional entropy of the jointly typical set when coding rate is larger than Shannon Capacity," Manuscript. 2006.

[9] X. Wu and L.-L. Xie, "AEP of output when rate is above capacity," in *Proc. of the 11th Canadian Workshop on Information Theory*, Ottawa, Canada, May 13-15, 2009.

[10] X. Wu and L.-L. Xie, "AEP of output when rate is above capacity: The Gaussian case" in *Proc. IEEE Int. Symposium on Information Theory*, Austin, Texas, USA, June 13-18, 2010.

[11] X. Wu and L.-L. Xie, "Asymptotic Equipartition Property of output when rate is above capacity," submitted to *IEEE Trans. Inform. Theory*, available online at http://arxiv.org/abs/0908.4445.

[12] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.