Forty-Fifth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 26-28, 2007

Coding Techniques for Primitive Relay Channels

Young-Han Kim

***Abstract*— We give a comprehensive discussion on coding techniques for the primitive relay channel, in which the channel input X is transmitted to the relay Y_1 and the ultimate receiver Y over a channel $p(y, y_1|x)$ and the relay can facilitate the communication between the transmitter and the receiver by sending some information to the receiver over a separate noiseless link of capacity R_0. In particular, we compare three known coding schemes, "decode-and-forward," "compress-and-forward," and "hash-and-forward," clarify their limitations, and explore further extensions. Possible unification of these coding schemes is also discussed, as well as a few open problems reflecting major difficulties in proving optimality.**

I. INTRODUCTION

The relay channel model was introduced by van der Meulen [23]. In this model, a relay node facilitates the communication between the transmitter X and the receiver Y by inferring about the transmitted signal X from its noisy observation Y_1 and by conveying this information with X_1.

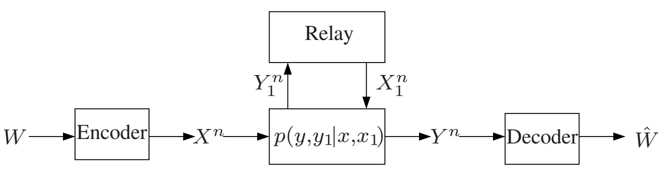

Fig. 1. The relay channel.

As the simplest single-source single-destination network model, the relay channel plays the role of a basic building block for future wireless networks and thus has received much attention recently. As shown in the definite work by Cover and El Gamal [7], previously known coding theorems provide a showcase of coding techniques in network information theory, combining 1) random codebook generation, 2) superposition coding, 3) successive cancellation, 4) Slepian–Wolf partitioning (random binning), 5) cooperative multiple access coding, 6) random covering, 7) list decoding, and 8) block Markov coding. Nonetheless, the capacity of the general relay channel is not known in general except for a few special cases such as degraded [7], semideterministic [12], orthogonal-component [15], and modulo-sum [2] relay channels. Recent articles by Kramer *et al.* [17] and El Gamal *et al.* [13] give nice surveys of the literature on single- or multiple-relay channels.

In wireless communication systems, a relay node cannot simultaneously transmit and receive signals in the same time or frequency band (full-duplex), it is natural to consider half-duplex models called *relay channels with orthogonal components* [18] as described in Figure 2. Half-duplex relaying

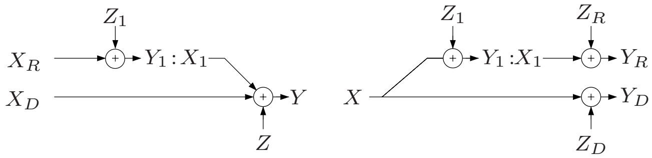

Fig. 2. Gaussian relay channels with orthogonal components.

decouples the transmitter–relay cooperation and the relay–receiver cooperation, which makes the capacity problem simpler than the general relay channel. In the first model (Figure 2, left), the transmitter and the relay communicate to the ultimate receiver over an additive white Gaussian multiple access channel $Y = X_D + X_1 + Z$. There is a separate Gaussian channel $Y_1 = X_R + Z_1$ from the transmitter to the relay, independent of the main multiple access channel. Whether the transmitter-to-relay link is noisy or not is irrelevant to the capacity of this relay channel, so we can replace the transmitter-to-relay link with a noiseless channel with the same capacity. This problem can be shown to be a special case of multiple access channel with partially cooperating encoders [24], the capacity region of which is known completely. More recently, the capacity for general orthogonal-component relay channels of the form $p(y, y_1|x_1, x) = p(y_1|x_R, x_1)p(y|x_D, x_1)$ has been established by El Gamal and Zahedi [15].

In the second model (Figure 2, right), the transmitter communicates to both the relay and the receiver over an additive white Gaussian noise broadcast channel ($Y_1 = X + Z_1,\ Y_D = X + Z_D$). The relay-to-receiver link is another Gaussian channel $Y_R = X + Z_R$, independent of the main broadcast channel. As in the first model, the relay-to-receiver channel can be replaced by a noiseless link of the same capacity. The capacity of this model is unknown in general, except for the degenerate case in which only the transmitter-to-relay link is very strong [17].

Motivated from the wireless relay channel in Figure 2 as well as hybrid wireless-wireline networks in which base stations are cooperating over high-speed optical communication links, this paper focuses on a rather limited class of relay channels, which we call *primitive relay channels*, as depicted in Figure 3. Here the channel input signal X is received by the relay Y_1 and the receiver Y through a channel $p(y, y_1|x)$, and the relay can communicate to the receiver over a separate noiseless link of rate R_0. We wish to communicate a message index $W \in [2^{nR}] = \{1, 2, \ldots, 2^{nR}\}$ reliably over this relay channel with a noiseless link. We specify a $(2^{nR}, 2^{nR_0}, n)$

The author is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 yhk@ucsd.edu

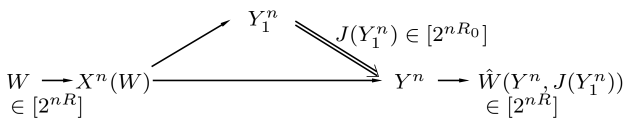

Fig. 3. Primitive relay channel.

code with an encoding function $X^n : [2^{nR}] \to \mathcal{X}^n$, a relay function $J : \mathcal{Y}_1^n \to [2^{nR_0}]$, and the decoding function $\hat{W} : \mathcal{Y}^n \times [2^{nR_0}] \to [2^{nR}]$. The probability of error is defined by $P_e^{(n)} = \Pr\{W \neq \hat{W}(Y^n, J(Y_1^n))\}$, with the message W distributed uniformly over $[2^{nR}]$. The capacity $C(R_0)$ is the supremum of all achievable rates R, i.e., the rates R for which there exists a sequence of $(2^{nR}, 2^{nR_0}, n)$ codes such that $P_e^{(n)}$ can be made to tend to zero as $n \to \infty$. Equivalently, we are interested in the achievable rate region, defined as the set of all (R, R_0) pairs such that there exists a sequence of $(2^{nR}, 2^{nR_0}, n)$ codes with $P_e^{(n)} \to 0$.

Compared to the general relay channel $p(y, y_1|x, x_1)$, the primitive relay channel model is much simpler, decoupling the multiple access component $p(y|x, x_1)$ from the broadcast component $p(y, y_1|x)$. Nonetheless, the primitive relay channel captures most essential features and challenges of relaying, and thus serves as a good testbed for new relay coding techniques. In a sense, the primitive relay channel is the simplest channel coding problem (from the transmitter's point of view) with a source coding constraint. At the same time, it is the simplest source coding problem (from the relay's point of view) for a channel code; the relay wishes to compress Y_1^n to help the receiver decode X^n. Naturally the following question takes the central attention for communication over the primitive relay channel:

> How should the relay Y_1 summarize its noisy observation of the intended signal X using R_0 bits?

To shed some light on this question, this paper reviews and clarifies three known coding techniques—decode-and-forward, compress-and-forward, and hash-and-forward. The main emphasis is on how these coding techniques compare with each other and how they can be improved and possibly unified. Several simple examples are presented to illustrate these points. It is hoped that this paper will bring more interests in this simple, yet fundamental problem of the primitive relay channel, leading to new coding and proof techniques.

The next section presents few known upper bounds on the capacity $C(R_0)$. In Section III, each coding technique along with the associated achievable rate is extensively discussed. Section IV concludes the paper with future directions and important open questions.

II. Outer Bounds on $C(R_0)$

For the general relay channel, the well-known cutset bound [9, Section 15.10] for multiterminal networks is given as follows:

$$C \leq \max_{p(x,x_1)} \min\{I(X, X_1; Y),\ I(X; Y, Y_1|X_1)\}. \tag{1}$$

For the primitive relay channel, it can be easily shown that the cutset bound (1) reduces to the following upper bound on the capacity (see Figure 4):

Proposition 1. *The capacity $C(R_0)$ of the primitive relay channel is upper bounded by*

$$C(R_0) \leq \max_{p(x)} \min\{I(X;Y) + R_0,\ I(X;Y,Y_1)\}. \tag{2}$$

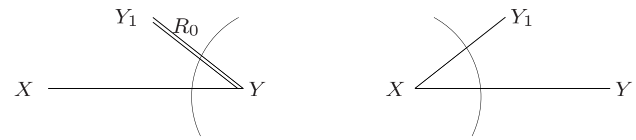

Fig. 4. Cutsets for the primitive relay channel.

Although it has been widely believed that the cutset bound is loose in general, finding a counterexample has turned out to be a nontrivial task. In fact, the bounds (1) and (2) are tight for almost all known capacity theorems for degraded [7], semideterministic [12], orthogonal-component [15], and semideterministic primitive [8] relay channels.

In [29], Zhang showed that

$$C(R_0) < C(0) + R_0 = \max_{p(x)} I(X;Y) + R_0$$

for

$$R_0 > \max_{p(x): I(X;Y) = C(0)} I(X;Y_1) - C(0),$$

provided that 1) Y and Y_1 are independent looks of X, i.e., $p(y, y_1|x) = p(y|x)p(y_1|x)$, and 2) Y is stochastically degraded from Y_1, i.e., there exists $q(y|y_1)$ such that $p(y|x) = \sum_{y_1} p(y_1|x) q(y|y_1)$. This result, in particular, shows that the cutset bound (2) can be loose with some positive gap.

A more direct example is constructed in a recent work by Aleksic *et al.* [2].

Example 1 (Modulo-sum relay)**.** Suppose the channel input X takes values in $\{0, 1\}$. Then X is observed at the receiver $Y = X + Z$ via a binary symmetric channel, and the channel noise Z is observed at the relay $Y_1 = Z + Z_1$ via another binary symmetric channel. More specifically, the channel is given by

$$\begin{aligned} Y &= X + Z \\ Y_1 &= Z + Z_1, \end{aligned}$$

where $Z \sim$ Bern$(1/2)$ and $Z_1 \sim$ Bern(δ), $0 < \delta < 1$, are independent of each other and X.

Using Mrs. Gerber's Lemma by Wyner and Ziv [27], we can show that

$$C(R_0) \leq 1 - H(\delta * H^{-1}(1 - R_0))$$

for $0 \leq R_0 \leq 1$. Here and henceforth, we use the notation $H(p)$ for a binary entropy function of a Bernoulli(p) random variable and $p * q = p(1-q) + q(1-p)$. From the compress-and-forward coding scheme we will discuss later, we can show that this rate is in fact achievable. In particular, we have $C(1-H(\delta)) = 1-H(\delta * \delta)$, which is strictly less than the cutset bound $1 - H(\delta)$. $\square$

For the details of the analysis along with extensions, refer to Aleksic *et al*. [2].

Incidentally, we can get a similar result by switching the roles of the receiver and the relay.

Example 2. Suppose the channel is given by

$$Y = Z + Z_1$$
$$Y_1 = X + Z,$$

where the input X takes binary values, and $Z \sim \text{Bern}(1/2)$ and $Z_1 \sim \text{Bern}(\delta)$, $0 < \delta < 1$ are independent of each other and X. Then, we have

$$C(R_0) = 1 - H(\delta * H^{-1}(1 - R_0))$$

for $0 \le R_0 \le 1$. $\square$

The above two examples are, however, very special in that under capacity-achieving input distribution $p^*(x)$, Y_1 is independent of Y. In general, the usual converse proof technique does not give a sharper bound than the cutset bound.

As a question simpler than fully characterizing $C(R_0)$, one may be interested in finding

$$R_0^* = \inf\{R_0 : C(R_0) = C(\infty) = \max_{p(x)} I(X;Y,Y_1)\}.$$

In other words, R_0^* is the smallest rate needed for the relay-to-receiver communication so that the maximum information rate $C(\infty) = \max_{p(x)} I(X;Y,Y_1)$ can be achieved. The cutset bound (2) leads to the following lower bound on R_0^*:

$$R_0^* \ge \min_{p(x): I(X;Y,Y_1) = C(\infty)} I(X;Y_1|Y).$$

Other than this cutset bound, the modulo-sum relay examples provide the only known (and tight) lower bound on R_0^*.

III. Lower Bounds on $C(R_0)$

A. Decode-and-Forward

We can easily see that a simple "multi-hop" scheme, in which the receiver fully decodes the message and forwards it with R_0 bits/symbol, can achieve

$$R_{\mathsf{MH}}(R_0) = \max_{p(x)} \min\{I(X;Y_1),\ R_0\}. \tag{3}$$

The major drawback of the multi-hop scheme, however, is that it does not use the receiver's own information about the message, inferred from the observation Y^n.

The "decode-and-forward" scheme [7, Section II] improves upon the multi-hop by incorporating the receiver Y's information about X with list decoding and random binning. More specifically, consider a randomly generated codebook of size 2^{nR} with each codeword $X^n(w)$, $w \in [2^{nR}]$ drawn independently from $p(x^n) = \prod_{i=1}^n p(x_i)$, and independently thrown into 2^{nR_0} bins. The entire codebook and the associated bin assignments are revealed to the transmitter, the relay, and the receiver prior to the actual communication. To communicate a message index w, the transmitter sends $X^n(w)$. Upon receiving Y^n, the receiver can form a list $L(Y^n)$ of jointly typical codewords $\{X^n(w)\}$ of size at most $2^{n(R-I(X;Y))+\epsilon}$ that contains the true codeword with high probability. On the other hand, if $R < I(X;Y_1)$, the relay can decode the correct message w with high probability and thus recover the associated bin index, which is sent to the receiver. Finally, if $R_0 > R - I(X;Y) + \epsilon$, i.e., if the number of bins is exponentially larger than the receiver's list size, then the receiver can correctly figure out the true codeword, which has the uniquely matching bin index with the one from the relay. Therefore, we get the following lower bound on $C(R_0)$.

Proposition 2. *For the primitive relay channel, we can achieve*

$$R_{\mathsf{DF}}(R_0) = \max_{p(x)} \min\{I(X;Y_1),\ I(X;Y) + R_0\}. \tag{4}$$

Compared to the multi-hop achievable rate (3), the achievable rate (4) has an additional term $I(X;Y)$, reflecting the information about the channel input X provided by the receiver's observation Y. On the other hand, compared to the cutset upper bound

$$C(R_0) \le \max_{p(x)} \min\{I(X;Y,Y_1),\ I(X;Y) + R_0\},$$

the lower bound (4) is tight when the relay is "close" to the transmitter, i.e.,

1) when the channel $p(y, y_1|x)$ is physically degraded as $X \to Y_1 \to Y$ and thus $I(X;Y_1) = I(X;Y,Y_1)$, or
2) $I(X;Y_1) \ge I(X;Y) + R_0$.

In general, the coding scheme requires the relay to decode the whole message and hence the transmitter-to-relay link becomes the major bottleneck of communication. The following example illustrates the point.

Example 3. Consider a primitive relay channel with input $X = (X_1, X_2)$, $X_1, X_2 \in \{0,1\}$, the relay output $Y_1 = X_1$ and the receiver output $Y = X_2$. Since X_2 cannot be decoded at the relay, it can be easily seen that

$$R_{\mathsf{DF}}(R_0) = R_0,$$

although $C(R_0) = R_0 + 1$ for $0 \le R_0 \le 1$. $\square$

To alleviate this problem, we can incorporate Cover's superposition coding [5] with the decode-and-forward coding scheme to get the so-called "partial decode-and-forward" scheme [7, Theorem 7]. Given $p(u, x)$ on a properly chosen auxiliary random variable U and the channel input X, we randomly generate a codebook consisting of $U^n(w_1) \sim \prod_{i=1}^n p(u_i)$, $w_1 \in [2^{nR_1}]$ and $X^n(w_2|w_1) \sim \prod_{i=1}^n p(x_i|U_i(w_1))$, $w_1 \in [2^{nR_1}], w_2 \in [2^{nR_2}]$. To send a message $w = (w_1, w_2) \in [2^{nR_1}] \times [2^{nR_2}]$ of rate $R = R_1 + R_2$, the transmitter sends $X^n(w_2|w_1)$. By operating the decode-and-forward scheme with the primitive relay channel $p(y, y_1|u) = \sum_x p(x|u)p(y, y_1|x)$ with pseudo-input U, we can achieve $R_1 = \min\{I(U;Y_1), I(U;Y) + R_0\}$. In addition, with correctly decoded $U^n(\hat{w}_1)$ the receiver can further deduce $X^n(w_2|\hat{w}_1)$ by joint typicality decoding if $R_2 < I(X;Y|U)$. Combining R_1 and R_2, we have the following lower bound on $C(R_0)$:

$$R_{\mathsf{PDF}}(R_0) = \max_{\substack{p(x,u) \\ |\mathcal{U}|\le|\mathcal{X}|}} \min\{I(U;Y_1)+I(X;Y|U), I(X;Y)+R_0\}. \quad (5)$$

Here the cardinality bound on U follows easily from the standard technique; refer to Wolfowitz [26, Section 12.2] and Salehi [21]. The partial decode-and-forward includes the original decode-and-forward; by taking $U = X$, we can readily see that (5) becomes (4).

Nonetheless, requiring the relay to decode something can be strictly suboptimal as demonstrated in the next example.

Example 4. Consider a primitive relay channel with the binary input $X \in \{0,1\}$, the receiver output $Y = X + Z$ through a binary symmetric channel with modulo-additive noise $Z \sim \text{Bern}(1/2)$ independent of X, and the relay output $Y_1 = Z$. In other words, the receiver gets a completely garbled X due to a random noise Z, while the receiver gets the noise itself. Because $Y_1 = Z$ is independent of X (and thus of U), the partial decode-and-forward can achieve no positive rate. In other words,

$$R_{\mathsf{PDF}}(R_0) = 0$$

for every $R_0 \ge 0$. But by describing Y_1 itself with 1 bit and using time-sharing, it can be easily checked that $C(R_0) = R_0$ for $0 \le R_0 \le 1$. □

B. Compress-and-Forward

While the relay in the decode-and-forward coding scheme strives to decode some digital information about X, the relay in the "compress-and-forward" (sometimes called "quantize-and-forward") coding scheme simply treats its signal Y_1^n as a random outcome generated by nature and strives to describe it as efficiently as possible. The details are as follows.

As in decode-and-forward, the codebook is generated at random under the input codebook $\{X^n(w)\}$ generated from $p(x)$. For the decoding of X^n, the relay and the receiver perform two stages of decoding operations. In the first stage, the relay and the receiver treat (Y_1^n, Y^n) as a source–side information pair, drawn as independent copies of $(Y, Y_1) \sim p(y, y_1) = \sum_x p(x)p(y, y_1|x)$. By covering Y_1^n space by $\hat{Y}_1^n$ compression codewords and using the Wyner–Ziv coding technique [28] for the source Y_1 with side information Y, the relay can convey $\hat{Y}_1^n$, jointly typical with Y_1^n for a given $p(\hat{y}_1|y_1)$, using $R_0 = I(Y_1; \hat{Y}_1|Y)$ bits/symbol. Note that at this stage, the receiver and the relay do not need to know the codebook, but only the input distribution $p(x)$.

In the second stage, the receiver decodes the codeword X^n using its own observation Y^n and the relay's description $\hat{Y}_1^n$. The Markov lemma [3] guarantees that the joint typicality decoding is correct with high probability if the code rate $R < I(X; Y, \hat{Y}_1)$. By choosing the best $p(x)$ and $p(\hat{y}_1|y_1)$, we have the following achievable rate:

Proposition 3. *For the primitive relay channel, we can achieve*

$$R_{\mathsf{CF}}(R_0) = \max_{\substack{p(x)p(\hat{y}_1|y_1) \\ |\hat{\mathcal{Y}}_1|\le|\mathcal{Y}_1|+1}} \{I(X;Y,\hat{Y}_1) : I(Y_1;\hat{Y}_1|Y) \le R_0\}. \quad (6)$$

Again the cardinality bound on $\hat{Y}_1$ follows from the standard arguments.

Processing the relay signal in blocks, compress-and-forward can be viewed as a natural extension of the "amplify-and-forward" coding scheme for Gaussian relay channels, in which the relay amplifies (or in general processes in a symbol-by-symbol manner) the observed signal. The achievable rate (6) is tight for a few special cases such as the modulo-sum relay channel [2] (cf. Examples 1 and 2) and the semideterministic case in which the relay output $Y_1 = f(X, Y)$ is a deterministic function of the input X and the receiver output Y [8].

There are, however, two obvious limitations of compress-and-forward. First, the achievable rate $R_{\mathsf{CF}}(R_0)$ is not a convex function in R_0 and thus can be improved by taking a convex envelope using time-sharing (cf. [14]) as

$$R_{\mathsf{CCF}}(R_0) = \max_{\substack{p(q)p(x|q)p(\hat{y}_1|y_1,q) \\ |\hat{\mathcal{Y}}_1|\le|\mathcal{Y}_1| \\ |\mathcal{Q}|\le 2}} \{I(X;Y,\hat{Y}_1|Q) : I(Y_1;\hat{Y}_1|Y,Q) \le R_0\}.$$

The second limitation comes from the fact that the relay in the compress-and-forward scheme does not use the codebook structure at all and simply treats Y_1^n as a random outcome. This can lead to a strictly suboptimal performance as illustrated by the next example.

Example 5. As in Example 3, the channel input $X = (X_1, X_2)$ consists of a pair of binary components X_1 and X_2. Suppose the relay observes the first component X_1 via a binary symmetric channel with crossover probability δ, i.e., $Y_1 = X_1 + Z$ where $Z \sim \text{Bern}(\delta)$ is independent of X. The receiver observes the second component X_2 without any error, i.e., $Y = X_2$.

By decoding X_1 and X_2 separately (partial decode-and-forward), we can easily achieve the capacity

$$C(R_0) = 1 + R_0$$

for $0 \le R_0 \le 1 - H(\delta)$. (The optimality follows from the cutset bound.) Hence, the minimum rate R_0^* necessary to achieve $C(\infty) = 2 - H(\delta)$ is $1 - H(\delta)$.

On the other hand, we have

$$\begin{aligned} R_{\mathsf{CF}}(R_0) &\le \max_{p(x)p(\hat{y}_1|y_1)} I(X;Y,\hat{Y}_1) \\ &\le \max_{p(x)} I(X;Y,Y_1) \\ &\le 2 - H(\delta), \end{aligned}$$

with equality only if X_1 and X_2 are independent and identically distributed according to $\text{Bern}(1/2)$. But under this choice of input distribution $p(x)$, we have

$$R = I(X;Y,\hat{Y}_1) = 2 - H(X_1|\hat{Y}_1),$$

while

$$R_0 = I(Y_1; \hat{Y}_1|Y) = 1 - H(Y_1|\hat{Y}_1).$$

Hence, from the conditional version of Mrs. Gerber's Lemma [27, Corollary 4] for the Markov chain $\hat{Y}_1 \to Y_1 \to X_1$, $R = 2 - H(\delta)$ implies that $R_0 = 1 > 1 - H(\delta)$. □

C. Hash-and-Forward

Suppose the relay output $Y_1 = f(X, Y)$ is a deterministic function of the channel input X and the receiver output Y. Given a randomly generated codebook under input distribution $p(x)$, the following decoding strategy achieves the cutset bound.

The receiver forms a list $L(Y^n)$ of $2^{n(R-I(X;Y)+\epsilon)}$ codewords that are jointly typical with given Y^n. Since Y_1 is a deterministic function of (X, Y), there is at most one Y_1^n to each pair of (X^n, Y^n). Therefore, given Y^n and the corresponding list $L(Y^n)$ of $2^{n(R-I(X;Y)+\epsilon)}$ codewords $X^n(w)$'s, the receiver can form a list of at most $2^{n(R-I(X;Y)+\epsilon)}$ Y_1^n's. Now the relay can use $R_0 = R - I(X;Y) + 2\epsilon$ bits/symbol to communicate the hash index of Y_1^n, which is sufficient to convey Y_1^n asymptotically error-free to the receiver. Finally, if $R < I(X; Y, Y_1)$ then with high probability there exists a unique codeword X^n in $L(Y^n)$ that is jointly typical with Y_1^n from the relay. (In other words, the correct codeword is uniquely jointly typical with (Y^n, Y_1^n).) Thus we can achieve

$$R_{\mathsf{HF}}(R_0) = \max_{p(x)} \min\{I(X;Y) + R_0,\ I(X;Y,Y_1)\},$$

which coincides with the cutset bound (2). The details of analysis and examples can be found in [8].

Although this coding scheme, called "hash-and-forward," uses list decoding at the receiver side and random binning to transfer the relay's information just as in the decode-and-forward scheme, the relay does not decode the message at all. In fact, the relay does not need to know the codebook or even $p(x)$.

On the other hand, the hash-and-forward is different from the compress-and-forward coding scheme described above in two aspects.

First, the relay output is transferred to the receiver by hashing only without any quantization with $\hat{Y}_1^n$, which would require $H(Y_1|Y)$ bits/symbol from the Slepian–Wolf coding theorem [22] if the compress-and-forward scheme was employed. Indeed, it is worthwhile to note that without quantization, the compress-and-forward achieves (R, R_0) pairs satisfying

$$R \le I(X; Y, Y_1) \tag{7a}$$
$$R_0 \ge H(Y_1|Y) \tag{7b}$$

given $p(x)$. Note that this region is achievable for any (not necessarily deterministic) primitive relay channel.

Second and more importantly, the hash-and-forward coding scheme can be viewed as another two-stage decoding method, the second stage being exactly the same as in compress-and-forward, while the first stage differs in that the receiver now exploits the codebook structure. In fact, the first stage of hash-and-forward can be reinterpreted as

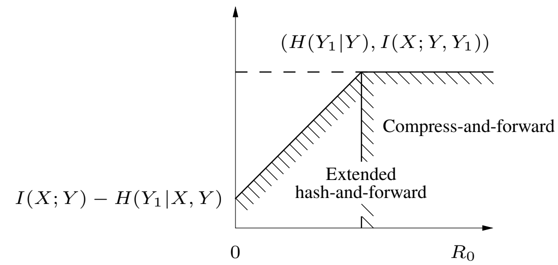

Fig. 5. The comparison of compress-and-forward achievable region (half-open rectangle) and hash-and-forward achievable achievable region (half-open trapezoid) when the relay transfers Y_1^n faithfully to the receiver.

follows. Upon receiving Y^n, the receiver forms a list $L(Y^n)$ of $2^{n(R-I(X;Y))}$ X^n codewords. Now using each (X^n, Y^n) pair as side information for describing Y_1^n by Slepian–Wolf coding, the relay needs to spend the communication rate of only $H(Y_1|X, Y) = 0$ since $Y_1 = f(X, Y)$. Since there are $2^{n(R-I(X;Y))}$ pairs of (X^n, Y^n), the relay can describe Y_1^n faithfully with rate $R_0 = R - I(X;Y)$.

The idea of employing the list decoding in the first stage of compress-and-forward by exploiting the codebook structure leads to a natural extension of hash-and-forward to a general nondeterministic primitive relay channels. Since the relay needs to spend the rate $H(Y_1|X, Y)$ for each (X^n, Y^n) pair and there are $2^{n(R-I(X;Y))}$ such pairs, the total rate required to transfer transfer Y_1^n becomes $R_0 = R - I(X;Y) + H(Y_1|X, Y)$. Hence for a general primitive relay channel, the "extended hash-and-forward" (or "codebook-aware compress-and-forward") coding scheme can achieve all (R, R_0) pairs satisfying

$$R \le I(X; Y, Y_1) \tag{8a}$$
$$R_0 \ge R - I(X;Y) + H(Y_1|X, Y) \tag{8b}$$

for some $p(x)$ (cf. (7)). Figure 5 compares the achievable rate regions described by (7) and (8) for a fixed $p(x)$. Although hash-and-forward seems to achieve more than compress-and-forward, the total achievable regions are in fact identical if we convexify the compress-and-forward region with no relaying $(R = I(X;Y), R_0 = 0)$ by time-sharing. We will comment more on this shortly.

We can further generalize the idea of codebook-aware compress-and-forward with random covering. Indeed, the relay requires the rate $I(Y_1; \hat{Y}_1|X, Y)$ to describe $\hat{Y}_1^n$ for each (X^n, Y^n) side information pair and there are $2^{n(R-I(X;Y))}$ pairs of (X^n, Y^n) that the receiver has to deal with, so $R_0 = R - I(X;Y) + I(Y_1; \hat{Y}_1|X, Y)$ suffices.

Proposition 4. *For the primitive relay channel, we can achieve*

$$R_{EHF}(R_0) = \max_{\substack{p(x)p(\hat{y}_1|y_1)\\ |\hat{\mathcal{Y}}_1| \le |\mathcal{Y}_1|+1}} \min\left\{ \begin{array}{c} I(X;Y,\hat{Y}_1), \\ I(X;Y) - I(Y_1;\hat{Y}_1|X,Y) + R_0 \end{array} \right\}. \tag{9}$$

Note that this achievable rate can be further extended by time-sharing. It is known [8] that for the semideter-

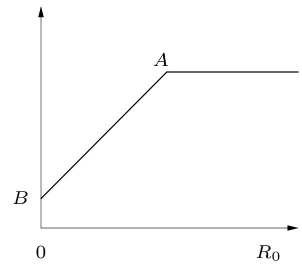

Fig. 6. The achievable rate region for the codebook-aware compress-and-forward coding scheme under a fixed $p(x)$.

ministic relay with $Y_1 = f(X,Y)$, both hash-and-forward and compress-and-forward achieve the capacity for each R_0. A natural question arises whether this coding technique of exploiting the codebook structure at the receiver side can strictly improve upon the original compress-and-forward scheme for a general primitive relay channel. The answer is somewhat disappointingly negative. To see this, consider the following diagram describing the achievable (R, R_0) pairs in (9) for a given $p(x)$. The corner point A in Figure 6, corresponding to

$$\begin{aligned} R &= I(X;Y,\hat{Y}_1) \\ R_0 &= R - I(X;Y) - I(Y_1;\hat{Y}_1|X,Y) = I(Y_1;\hat{Y}_1|Y), \end{aligned}$$

can be achieved also by compress-and-forward; see Proposition 3. On the other hand, the corner point B,

$$\begin{aligned} R &= I(X;Y) - I(Y_1;\hat{Y}_1|X,Y) \le I(X;Y) \\ R_0 &= 0, \end{aligned}$$

can be again achieved by not using the relay at all (compress-and-forward with $\hat{Y}_1 = 0$). Therefore, by time-sharing these two operating points, the convexified compress-and-forward coding scheme can achieve the entire convex envelope of $R_{\mathsf{EHF}}(R_0)$. (The other direction of inclusion is obvious.) This implies that there is no performance improvement if only the receiver uses the codebook structure while the relay still treats Y_1^n as a random outcome generated by nature.

IV. Discussion

Decode-and-forward and compress-and-forward represent two extreme methods of summarizing the relay's noisy observation about the message. In a sense, decode-and-forward extracts the digital information by decoding, while compress-and-forward processes the analog information by vector quantization with side information. Hash-and-forward, seemingly augmenting the quantization process with the receiver's knowledge about the digital structure, does not gain from the vanilla compress-and-forward.

That neither of decode-and-forward and compress-and-forward dominates the other, not to mention neither achieves the capacity in general, hints that optimal coding for a general primitive relay channel would require a completely new technique. To be fair, we can overlay decode-and-forward and compress-and-forward together to obtain a super coding scheme (see [7, Theorem 7] and [4]), but there is a strong indication that this super coding scheme can be strictly suboptimal [14].

The major conceptual question towards a new relay coding technique would be "How can the relay exploit the codebook structure?". In a sense, (partial) decode-and-forward can be viewed as compress-and-forward in which the relay uses the codebook $X^n(w)$ (or $U^n(w_1)$ in partial decode-and-forward) instead of $\hat{Y}_1^n$ as the covering code. Can we extend this idea to the case in which the relay cannot decode the codeword and instead transfer a list of codewords that are compatible with Y_1^n? This problem seems to require a novel combinatorial method on joint typicality graphs.

On the other hand, there are channel models in which it is impossible for the relay to exploit the codebook structure and thus the vanilla compress-and-forward scheme seems to be naturally optimal. As a first example, we consider a Gaussian channel with state.

Question 1. Let the channel output Y is given by $X + Z + S$ where the input X has average power constraint P, the additive Gaussian noise Z has variance N, and the additional additive Gaussian interference S has variance Q. We assume that Z and S are independent of each other and of X. The capacity of this channel (without any relay) is

$$C = C(0) = \frac{1}{2}\log\left(1 + \frac{P}{N+Q}\right).$$

Now suppose the relay observes $Y_1 = S$ and tries to help the receiver by sending some information at rate R_0. What is the capacity $C(R_0)$? □

When $N = 0$, it is known [8] that $C(R_0) = C(0) + R_0$ for every R_0. The answer is not known in general.

Ahlswede and Han [1, Section V] considered a general state-dependent channel $p(y|x,s)$ with rate-limited state information available at the receiver. By identifying $Y_1 = S$ with the channel state S independent of the channel input X, this model can be regarded as a special case of primitive relay channels. Although compress-and-forward is strongly believed to be optimal as Ahlswede and Han conjectured, the proof seems to be out of reach at the moment except for a few special cases considered in [8], [2].

The next example reveals an interesting connection between the primitive relay channel and coding for distributed computing.

Question 2. Suppose a binary input X is observed at the receiver and the relay through two identical and *dependent* erasure channels. More specifically,

$$(Y, Y_1) = \begin{cases} (X, X), & \text{with probability } 1/3 \\ (X, e), & \text{with probability } 1/3 \\ (e, X), & \text{with probability } 1/3, \end{cases}$$

where e denotes the erasure symbol. One salient feature of this primitive relay channel is that X is a deterministic function of (Y, Y_1). Hence, if the receiver is fully aware of Y_1 as well as Y, it can determine X without any error and achieve the 1-bit capacity. Obviously, using $R_0 = H(Y_1) = \log 3$ bits/symbol, the relay can describe Y_1 faithfully. Or by using Slepian–Wolf coding, it suffices to spend $R_0 = H(Y_1|Y) = 1$. What is the minimum R_0 to achieve $C(R_0) = 1$, or in our notation, what is R_0^*? $\square$

In [20], Orlitsky and Roche considered the following interesting problem. Suppose there is a pair of memoryless sources (U^n, V^n). Alice observes U^n and Bob observes V^n. How many bits would Alice need to describe U^n so that Bob can calculate some symbol-by-symbol function $f(U^n, V^n) = (f(U_1, V_1), \dots, f(U_n, V_n))$? This problem can be viewed as a special case of the Wyner–Ziv problem with side-information-dependent distortion measure [10, Corollary 3.4.6] with side information V and zero Hamming distortion between the reconstruction $\hat{U}$ and the desired function value $f(U, V)$. However, Orlitsky and Roche went further to show that 1) this quantity is identical to the conditional graph entropy $H_G(U|V)$ of the characteristic graph G of (U, V) and f defined by Witsenhausen [25], and 2) Bob can calculate the desired function with asymptotically vanishing block error probability, instead of symbol error probability in the usual Wyner–Ziv setup. See Doshi *et al.* [11] for a recent development.

Now turning back to our erasure relay problem, evaluating $H_G(Y_1|Y)$ under $X \sim \text{Bern}(1/2)$ shows that

$$R_0 = \frac{2}{3} H\left(\frac{1}{4}\right),$$

which is less than $H(Y_1|Y) = 1$, is sufficient to achieve $C = 1$. Can we prove the optimality of this rate? In general, is $R_0^* = H_G(Y_1|Y)$ for deterministic primitive relay channels with $X = f(Y, Y_1)$? The standard weak converse proof techniques seem to fail.

We conclude this paper with another example with two relays, which shows a connection to helper problems [10, Section 3.1] and thus hints further difficulties in finding an optimal coding technique.

Example 6. Suppose the receiver $Y = X + Z$ observes the binary input X over a binary symmetric channel with crossover probability δ. There are two relays; the first relay receives $Y_1 \sim \text{Bern}(1/2)$, independent of X, and the second relay receives $Y_2 = Y_1 + Z$, again independent of X. How many bits R_1 and R_2 should the relays respectively communicate to the receiver to achieve $C = 1$ bit?

Under Slepian–Wolf coding for (Y_1, Y_2) with side information Y, which can be easily proved to be the optimal compress-and-forward scheme to achieve $C = 1$, we need $R_1 + R_2 = 1 + H(\delta)$. However, using the Körner–Marton encoding [16] of binary doubly symmetric sources, $R_1 = R_2 = H(\delta)$ is sufficient to achieve $C = 1$ and in fact optimal, which can be shown from the cutset bound. $\square$

References

[1] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 396–412, 1983.

[2] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channel," submitted to *IEEE Int. Symp. Inf. Theory,* Nice, France, 2007. [Online]. Available: http://arxiv.org/abs/0704.3591/

[3] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978.

[4] H.-F. Chong, M. Motani, and H. K. Garg, "Generalized backward decoding strategies for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-53, no. 1, pp. 394–401, 2007.

[5] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[6] ——, "The capacity of the relay channel," in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, Eds. New York: Springer-Verlag, 1987, pp. 72–73.

[7] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 5, pp. 572–584, Sept. 1979.

[8] T. M. Cover and Y.-H. Kim, "Capacity of a class of determinstic relay channels," submitted to *IEEE Trans. Inf. Theory,* November 2006. [Online]. Available: http://arxiv.org/abs/cs.IT/0611053/

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[10] I. Csiszár and J. Körner, *Information Theory*, 3rd ed. Budapest: Akadémiai Kiadó, 1981.

[11] V. Doshi, D. Shah, M. Medard, and S. Jaggi, "Graph coloring and conditional graph entropy," in *Proc. Fortieth Asilomar Conference Signals, Syst. Comp.*, Monterrey, CA, Oct./Nov. 2006, pp. 2137–2141.

[12] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 3, p. 536, May 1982.

[13] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delay," submitted to *IEEE Trans. Inf. Theory,* 2006.

[14] 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.

[15] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inform. Theory*, vol. IT-51, no. 5, pp. 1815–1817, 2005.

[16] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 219–221, 1979.

[17] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. IT-51, no. 9, pp. 3037–3063, Sept. 2005.

[18] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. IT-50, no. 12, pp. 3062–3080, 2004.

[19] One Laptop per Child Project. [Online]. Available: http://www.onelaptop.org/

[20] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inform. Theory*, vol. IT-47, no. 3, pp. 903–917, 2001.

[21] M. Salehi, "Cardinality bounds on auxiliary variables in multiple-user theory via the method of Ahlswede and K orner," Department of Statistics, Stanford University, Technical Report 33, Aug. 1978.

[22] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, 1973.

[23] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[24] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 441–445, 1983.

[25] H. S. Witsenhausen, "Values and bounds for the common information of two discrete random variables," *SIAM J. Appl. Math.*, vol. 31, no. 2, pp. 313–333, 1976.

[26] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Berlin: Springer-Verlag, 1978.

[27] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications—I," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 769–772, 1973.

[28] ——, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, 1976.

[29] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inform. Theory*, vol. IT-34, no. 5, pp. 1106–1110, Sept. 1988.