On the Role of Estimate-and-Forward With Time Sharing in Cooperative Communication

Ron Dabora and Sergio D. Servetto

Abstract—**In this paper, we focus on the general relay channel. We investigate the application of the estimate-and-forward (EAF) relaying scheme to different scenarios. Specifically, we study assignments of the auxiliary random variable that always satisfy the feasibility constraints. We then consider the Gaussian relay channel with coded modulation, where we show that a three-level quantization outperforms the Gaussian quantization commonly used to evaluate the achievable EAF rates in this scenario. Last, we consider the cooperative general broadcast scenario with a multistep conference between the receivers. We first apply EAF to obtain a general achievable rate region with a multistep conference. We then use an explicit assignment for the auxiliary random variables to obtain an explicit rate expression for the single common message case with a two-step conference.**

Index Terms—**Channel capacity, cooperative broadcast, estimate-and-forward (EAF), network information theory, relay channels.**

I. Introduction

The relay channel was introduced by van der Meulen in 1971 [1]. In this setup, a single transmitter with channel input X^n communicates with a single receiver with channel output Y^n, where the superscript n denotes the length of a vector. In addition, an external transceiver, called a relay, listens to the channel and can input signals to the channel. We denote the channel output at the relay with Y_1^n and its channel input with X_1^n. This setup is depicted in Fig. 1.

A. Relaying Strategies

In [2] Cover and El-Gamal introduced two relaying strategies commonly referred to as decode-and-forward (DAF) and estimate-and-forward (EAF). In DAF, the relay decodes the message sent from the transmitter and then, at the next time interval, transmits a codeword based on the decoded message. The rate achievable with DAF is given in [2, Th. 1] stated below.

Manuscript received October 30, 2006; revised October 23, 2007. Current version published September 17, 2008. This work was supported by the National Science Foundation under Awards CCR-0238271 (CAREER), CCR-0330059, and ANR-0325556. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Nice, France, June 2007.

R. Dabora was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (e-mail: rd97@cornell.edu).

S. D. Servetto, deceased, was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: servetto@ece.cornell.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Color versions of Figures 3–7 and 9 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.928981

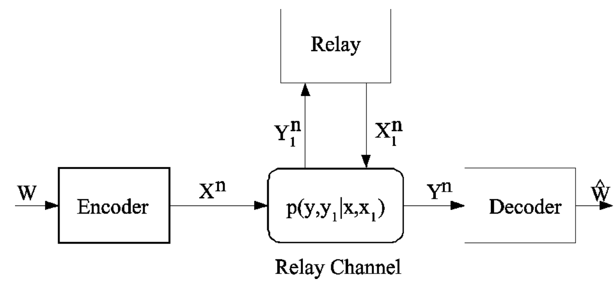

Fig. 1. Relay channel. The encoder sends a message W to the decoder.

Theorem 1 (Achievability of [2, Th. 1]): For the general relay channel, any rate R satisfying

$$R \leq \min\{I(X, X_1; Y), I(X; Y_1|X_1)\} \tag{1}$$

for some joint distribution $p(x, x_1, y, y_1) = p(x, x_1)p(y, y_1|x, x_1)$, is achievable.

We note that for DAF to be effective, the rate to the relay has to be greater than the point-to-point rate, i.e.,

$$I(X; Y_1|X_1) > I(X; Y|X_1) \tag{2}$$

otherwise higher rates could be obtained without using the relay at all. For relay channels where DAF is not useful or not optimal, Cover and El-Gamal proposed the EAF strategy. In this strategy, the relay sends an estimate of its channel output to the destination, without decoding the source message at all. The achievable rate with EAF is given in [2, Th. 6] stated below.

Theorem 2 [2, Th. 6]: For the general relay channel, any rate R satisfying

$$R \leq I(X; Y, \hat{Y}_1|X_1) \tag{3}$$

$$\text{subject to } I(X_1; Y) \geq I(Y_1; \hat{Y}_1|X_1, Y) \tag{4}$$

for some joint distribution $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1)$, $\hat{\mathcal{Y}}_1$ is an arbitrary sample space, is achievable.

We note that a time-sharing (TS) random variable can improve the achievable rate with EAF [18, Th. 2].

Of course, one can combine the DAF and EAF schemes by performing partial decoding at the relay, thus obtaining higher rates as in [2, Th. 7]. We note that the name "estimate-and-forward" comes from the original description of Cover and El-Gamal in [2, Sec. VI], where $\hat{Y}_1$ is referred to as an *estimate*. Another common name for this strategy is compress-and-forward (CAF, also sometimes abbreviated as CF), and it is also referred to by the names observe-and-forward and quantize-and-forward; see [8, Sec. I-B]. In the following, we will refer to this method as EAF. Also, DAF is sometimes abbreviated as DF.

B. Related Work

In recent years, the research in relaying has mainly focused on multiple-level relaying and the multiple-input–multiple-output (MIMO) relay channel. In the context of multiple-level relaying based on DAF, several DAF variations were considered. In [3], Cover and El-Gamal's block Markov encoding/successive decoding DAF method was applied to the multiple-relay case. Later work [4]–[6] applied the so-called regular encoding/sliding-window decoding and the regular encoding/backward decoding techniques to the multiple-relay scenario. In [7], the DAF strategy was applied to the MIMO relay channel. The EAF strategy was also applied to the multiple-relay scenario. The work in [8], for example, considered the EAF strategy for multiple-relay scenarios and the Gaussian relay channel, in addition to considering the DAF strategy. The EAF strategy in the multiple-relay setup was also considered in [9]. Another approach applied recently to the relay channel is that of iterative decoding. In [10], the three-node network in the half-duplex regime was considered. In the relay case, [10] uses an iterative scheme where the receiver first uses EAF to send information to the relay and then the relay decodes and uses DAF at the next time interval to help the receiver. Combinations of EAF and DAF were also considered in [11], where conferencing schemes over orthogonal relay–receiver channels were analyzed and compared. Both [10] and [11] focus on the Gaussian case. EAF was also applied to tree networks in [12], where the links between the nodes in the network (except the source node) were assumed orthogonal.

An extension of the relay scenario to a hybrid broadcast/relay system was introduced in [13] in which the authors applied a combination of an EAF step followed by a DAF step to the independent broadcast channel (BC) with a single common message, and then extended this scheme to a multistep conference. In [14], we used both a single-step and a two-step conference with orthogonal conferencing channels in the discrete memoryless framework. A thorough investigation of the broadcast–relay channel was carried out in [15], where the authors applied the DAF strategy to the case where only one user is helping the other user, and also presented an upper bound for this case. Then, the fully cooperative scenario was analyzed. The authors applied both the DAF and the EAF methods to that case. Finally, we note the work in [16] where both transmitter cooperation and receiver cooperation were considered (separately) over wireless (i.e., correlated) cooperation channels, with a single cooperation cycle. In transmitter cooperation, DAF in combination with dirty-paper coding (see [17]) was considered and for receiver cooperation EAF and DAF were considered, focusing on Gaussian channels.

C. The Gaussian Relay Channel With Coded Modulation

One important instance of the relay channel we consider in this work is the Gaussian relay channel with coded modulation. This scenario is important in evaluating the rates achievable with practical communication systems, where components in the receive chain, such as equalization, for example, require a uniformly distributed finite constellation for optimal operation. In Gaussian relay channel scenarios, the following three types for relaying schemes are most frequently encountered.

- The DAF scheme. This scheme achieves capacity for the physically degraded Gaussian relay channel (see [2, Sec. IV]). In [8, Sec. VII-B], it is shown that for asymptotically high signal-to-noise ratio (SNR) on the source-relay link, DAF achieves the capacity of the Gaussian relay channel.
- The EAF scheme in which the auxiliary variable $\hat{Y}_1|Y_1$ is assigned a Gaussian distribution. For example, in [18, Sec. IV], a Gaussian auxiliary random variable (RV) is used together with TS at the transmitter. The Gaussian assignment achieves capacity for the Gaussian relay channel when the SNR on the relay-destination link approaches infinity [8, Remark 31]. In [19], an achievable rate with full duplex relay transmission employing Gaussian EAF over the Rayleigh relay channel is obtained for the high SNR regime.[1]
- The linear relaying scheme in which the relay transmits a weighted sum of all its previously received channel outputs [18, Sec. V]. Amplify-and-forward (AAF) is an important subclass of this family of relaying functions. In [20], AAF was combined with DAF resulting in the decode-amplify-and-forward scheme. A related approach to AAF was proposed in [21], in which the relay finds a minimum mean squared error estimate of its received symbol on a symbol-by-symbol basis, and uses it to generate its transmitted symbol.

Several recent papers consider the Gaussian relay channel with coded modulation. In [22], the author considered the performance of half-duplex DAF relaying for different practical systems. In [23], DAF and AAF were considered for coherent binary phase-shift keying (BPSK) and, in [21], examples with BPSK were considered as well. In [24], the optimal uncoded regeneration function for BPSK in the two-hop setup, for minimum average probability of error at the destination, was derived.

As indicated by several authors (see [18]), it is not obvious if a Gaussian relay mapping is indeed optimal. In this paper, we show that for the case of coded modulation, there are scenarios where non-Gaussian assignments of the auxiliary RV result in a higher rate than the commonly applied Gaussian assignment.

D. Main Contributions

In the following, we summarize the main contributions of this work.

- We first present an alternative characterization for the EAF rate of [2, Th. 6] that does not have a feasibility constraint. This result is derived via an assignment of the auxiliary RV we call the TS assignment. Thus, this characterization provides a positive rate for any auxiliary map-

[1] In [19], Y_1 is not Gaussian but compression is performed by adding a Gaussian RV to the received signal at the relay. Knowledge of the fading coefficients is assumed. High SNR is assumed for both the relay and the receiver.

ping $p(\hat{y}_1|x_1, y_1)$, as long as the channel setup permits a positive rate. We also consider the achievable rate for the single-relay channel when the destination receiver uses joint decoding of both the relay information and the source message, instead of the sequential decoding used in [2, Th. 6]. The rate expression can be obtained as a special case of [26] by not performing partial decoding at the relay. We then show that joint decoding does not increase the maximum rate of the EAF strategy, and find an assignment for the auxiliary RV in [2, Th. 6] that obtains the joint-decoding rate expression from the general EAF expression. We also present another assignment that results in a rate that is always at least as high as the joint-decoding rate, for the same distribution chain.

- We consider the optimization of the EAF mapping for the Gaussian relay channel with an orthogonal relay-destination link. We focus on the coded modulation scenario with BPSK modulation at the transmitter, and show that there are three regions: high SNR on the source–relay link, where DAF is the best strategy (out of DAF, Gaussian EAF, and the three-level quantization we propose), low SNR on the source–relay link in which the common EAF with Gaussian assignment is best, and an intermediate region where EAF with "hard decision (HD) per symbol" is better than both DAF and Gaussian EAF. For this intermediate SNR region, we study for the first time two types of HDs: deterministic and probabilistic, and show that each one of them can be superior, depending on the channel conditions. In comparison, previous work in which EAF was applied to this scenario used the Gaussian auxiliary mapping, and did not consider optimization of the mapping.
- Last, we consider the cooperative broadcast scenario with receivers holding a multistep conference. We present a general rate region, extending the Marton rate region of [25, Th. 2] to the case where the receivers hold a K-cycle conference prior to decoding the messages. The conference is based on successive EAF steps. This improves on previous results by letting the receivers successively refine the conference information, thus potentially allowing them to achieve higher source-destination rates. We then specialize this result to the single common message case and obtain explicit expressions (without auxiliary RVs) for the two-step conference.

The rest of this paper is organized as follows. In Section II, we discuss the TS-EAF assignment. We consider the EAF strategy with TS and relate it to the EAF rate expression with joint decoding at the destination receiver. In Section III, we study the Gaussian relay channel with coded modulation. In Section IV, we consider the general cooperative broadcast scenario and obtain an explicit rate expression by applying TS-EAF to the multistep conference. Finally, Section V presents concluding remarks.

II. Time-Sharing Estimate-and-Forward

A. Definitions

First, a word about notation: we denote random variables with capital letters, e.g., X and Y, and their realizations with lower case letters x and y. A random variable X takes values in a set $\mathcal{X}$. We use $||\mathcal{X}||$ to denote the cardinality of a finite, discrete set $\mathcal{X}$, and $p_X(x)$ to denote the probability distribution function (pdf) of X on $\mathcal{X}$. For brevity, we may omit the subscript X when it is obvious from the context. We denote vectors with boldface letters, e.g., $\mathbf{x}$ and $\mathbf{y}$; the ith element of a vector $\mathbf{x}$ is denoted by x_i and we use x_i^j where $i < j$ to denote $(x_i, x_{i+1}, \dots, x_{j-1}, x_j)$. Unless otherwise specified, we set $\mathbf{x} \equiv x^n \equiv x_1^n$. We use $A_\epsilon^{*(n)}(X)$ to denote the set of ϵ-strongly typical sequences with respect to (w.r.t.) distribution $p_X(x)$ on $\mathcal{X}$, as defined in [27, Ch. 5.1].[2] Finally, we denote by $\mathcal{N}(a, b)$ a Gaussian RV with mean a and variance b. We also have the following definitions.

Definition 1: The discrete relay channel is defined by two discrete input alphabets $\mathcal{X}$ and $\mathcal{X}_1$, two discrete output alphabets $\mathcal{Y}$ and $\mathcal{Y}_1$, and a collection of pdfs $p(y, y_1|x, x_1)$ giving the probability distribution on $\mathcal{Y} \times \mathcal{Y}_1$ for each $(x, x_1) \in \mathcal{X} \times \mathcal{X}_1$. The relay channel is called *memoryless* if the probability of the outputs at time i satisfies

$$p\left(y_i, y_{1,i}|x^i, x_{1,1}^i, y^{i-1}, y_{1,1}^{i-1}, w\right) = p\left(y_i, y_{1,i}|x_i, x_{1,i}\right)$$

where w is the transmitted message.

In this paper we consider only the memoryless relay channel.

Definition 2: A $(2^{nR}, n)$ *code* for the relay channel consists of a source message set $\mathcal{W} = \left\{1, 2, \dots, 2^{nR}\right\}$, a mapping function f at the encoder

$$f : \mathcal{W} \mapsto \mathcal{X}^n$$

a set of n relay functions

$$x_{1,i} = t_i\left(y_{1,1}, y_{1,2}, \dots, y_{1,i-1}\right), \qquad i > 1$$

where the ith relay function t_i maps the first $i-1$ channel outputs at the relay into a transmitted relay symbol at time i, and $x_{1,1} = a$, $a \in \mathcal{X}_1$, is an arbitrary constant. Last, we have a decoder

$$g : \mathcal{Y}^n \mapsto \mathcal{W}.$$

Definition 3: The average probability of error for a code of length n for the relay channel is defined as

$$P_e^{(n)} = \Pr(g(Y^n) \neq W)$$

where W is selected uniformly over $\mathcal{W}$.

Definition 4: A rate R is called *achievable* if there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$.

B. Single-Relay EAF With TS

Consider the following assignment of the auxiliary random variable of Theorem 2:

$$p(\hat{y}_1|y_1, x_1) = \begin{cases} q, & \hat{y}_1 = y_1 \\ 1-q, & \hat{y}_1 = E \notin \mathcal{Y}_1 \end{cases} \tag{5}$$

[2]Let $N(x^n|a) \triangleq$ number of times the symbol $a \in \mathcal{X}$ appears in the sequence $x^n \in \mathcal{X}^n$. The strongly ϵ-typical set w.r.t. probability distribution $p_X(x)$ is defined as the set of all sequences $x^n \in \mathcal{X}^n$ such that $N(x^n|a) = 0$ if $p_X(a) = 0$ and $\sum_{a\in\mathcal{X}} \left|\frac{1}{n}N(x^n|a) - p_X(a)\right| \leq \epsilon$. We denote this set with $A_\epsilon^{*(n)}(X)$.

where "E" can be viewed as an erasure symbol. Under this assignment, the feasibility condition of (4) becomes

$$\begin{aligned} I(X_1;Y) \geq\ & I(Y_1;\hat{Y}_1|X_1,Y) \\ =\ & H(Y_1|X_1,Y) - (1-q)H(Y_1|X_1,Y) \\ & - qH(Y_1|X_1,Y,Y_1) \\ =\ & qH(Y_1|X_1,Y) \end{aligned}$$

and the rate expression (3) becomes

$$\begin{aligned} R &\leq I(X;Y,\hat{Y}_1|X_1) \\ &= I(X;Y|X_1) + H(X|X_1,Y) \\ &\quad - (1-q)H(X|X_1,Y) - qH(X|X_1,Y,Y_1) \\ &= I(X;Y|X_1) + qI(X;Y_1|X_1,Y). \end{aligned}$$

Clearly, maximizing the rate implies maximizing q subject to the constraint $q \in [0,1]$. This gives the following corollary to Theorem 2.

Corollary 1: For the general relay channel, any rate R satisfying

$$R \leq I(X;Y|X_1) + \left[\frac{I(X_1;Y)}{H(Y_1|X_1,Y)}\right]^* I(X;Y_1|X_1,Y) \quad (6)$$

for the joint distribution $p(x,x_1,y,y_1) = p(x)p(x_1)p(y,y_1|x,x_1)$, with $[x]^* \triangleq \min(x,1)$, is achievable.

Now, consider the following distribution chain:

$$p(x,x_1,y,y_1,\hat{y}_1,\hat{\hat{y}}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)p(\hat{\hat{y}}_1|\hat{y}_1). \quad (7)$$

We note that this extended chain can be put into the standard form by letting $p(\hat{\hat{y}}_1|x_1,y_1) = \sum_{\hat{y}_1} p(\hat{y}_1,\hat{\hat{y}}_1|x_1,y_1) = \sum_{\hat{y}_1} p(\hat{y}_1|x_1,y_1)p(\hat{\hat{y}}_1|\hat{y}_1)$. After compression of Y_1 into $\hat{Y}_1$, there is a second compression operation, compressing $\hat{Y}_1$ into $\hat{\hat{Y}}_1$. The output of the second compression is used to facilitate cooperation between the relay and the destination. Therefore, the receiver decodes the message based on $\hat{\hat{\mathbf{y}}}_1$ and $\mathbf{y}$, using exactly the same steps as in the standard relay decoding, with $\hat{\hat{\mathbf{y}}}_1$ replacing $\hat{\mathbf{y}}_1$. Then, the expressions of Theorem 2 become

$$R \leq I(X;Y,\hat{\hat{Y}}_1|X_1) \quad (8)$$

$$\text{subject to } I(X_1;Y) \geq I(Y_1;\hat{\hat{Y}}_1|X_1,Y). \quad (9)$$

Now, applying TS to $\hat{\hat{Y}}_1$ with

$$p(\hat{\hat{y}}_1|\hat{y}_1) = \begin{cases} q, & \hat{\hat{y}}_1 = \hat{y}_1 \\ 1-q, & \hat{\hat{y}}_1 = E \notin \hat{\mathcal{Y}}_1 \end{cases} \quad (10)$$

the expressions in (8) and (9) become

$$\begin{aligned} R &\leq I(X;Y|X_1) + I(X;\hat{\hat{Y}}_1|X_1,Y) \\ &= I(X;Y|X_1) + qI(X;\hat{Y}_1|X_1,Y) \quad (11) \\ I(X_1;Y) &\geq I(Y_1;\hat{\hat{Y}}_1|X_1,Y) \\ &= qI(Y_1;\hat{Y}_1|X_1,Y). \quad (12) \end{aligned}$$

Combining this with the constraint $q \in [0,1]$, we obtain the following proposition.

Proposition 1: For the general relay channel, any rate R satisfying

$$R \leq I(X;Y|X_1) + \left[\frac{I(X_1;Y)}{I(Y_1;\hat{Y}_1|X_1,Y)}\right]^* I(X;\hat{Y}_1|X_1,Y)$$

for some joint distribution $p(x,x_1,y,y_1,\hat{y}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)$, is achievable.

This proposition generalizes on Corollary 1 by performing a general Wyner–Ziv (WZ) compression combined with TS (which is a specific type of WZ compression), intended to guarantee feasibility of the first compression step. Comparing with [2, Th. 6], we see that for the same distribution chain, Proposition 1 always provides a positive rate (as long as the channel permits). In Section III, we apply a similar idea to EAF relaying in the Gaussian relay channel with coded modulation. We note that taking the supremum over all mappings, we get that Proposition 1 achieves the same rate as Theorem 2; see Appendix A.

C. Joint Decoding and Time Sharing

In the original work of [2, Th. 6], the decoding procedure at the destination receiver for decoding the message w_{i-1} at time i consists of three steps (the notations below are identical to [2, Th. 6]; the reader is referred to the proof of [2, Th. 6] to recall the definitions of the sets and variables used in the following description).

1) Decode the relay index s_i using $\mathbf{y}(i)$, the received signal at time i.
2) Decode the relay message z_{i-1}, using s_i, the received signal $\mathbf{y}(i-1)$, and the previously decoded s_{i-1}.
3) Decode the source message w_{i-1} using $\mathbf{y}(i-1)$, z_{i-1}, and s_{i-1}.

Evidently, when decoding the relay message z_{i-1} at the second step, the receiver does not make use of the statistical dependence between $\hat{\mathbf{y}}_1(i-1)$, the relay sequence at time $i-1$, and $\mathbf{x}(w_{i-1})$, the transmitted source codeword at time $i-1$. The way to use this dependence is to jointly decode z_{i-1} and w_{i-1} after decoding s_i and s_{i-1}. The joint-decoding procedure then has the following steps.

1) From $\mathbf{y}(i)$, the received signal at time i, the receiver decodes s_i by looking for a unique $s \in \mathcal{S}$, the set of indices used to enumerate the $\mathbf{x}_1$ sequences, such that $(\mathbf{x}_1(s),\mathbf{y}(i)) \in A_\epsilon^{*(n)}$. As in [2, Th. 6], the correct s_i can be decoded with an arbitrarily small probability of error by taking n large enough as long as
$$R_0 \leq I(X_1;Y) \quad (13)$$
where $||\mathcal{S}|| = 2^{nR_0}$.
2) The receiver now knows the set S_{s_i} into which z_{i-1} (the relay message at time $i-1$) belongs. Additionally, from decoding at time $i-1$, the receiver knows s_{i-1}, used to generate z_{i-1}. The receiver generates the set $\mathcal{L}(i-1) = \left\{w \in \mathcal{W} : (\mathbf{x}(w),\mathbf{y}(i-1),\mathbf{x}_1(s_{i-1})) \in A_\epsilon^{*(n)}\right\}$. Assuming no decoding error at the previous step, the average

size of this set (averaged over all selections of codewords $\mathbf{x}_1$ and received sequences $\mathbf{y}$) is

$$E\left\{ ||\mathcal{L}(i-1)|| \Big| \text{ no error at time } i-1 \right\} \leq 1 + 2^{n(R-I(X;Y|X_1)+\eta)}.$$

3) The receiver now looks for a unique $w \in \mathcal{L}(i-1)$ such that $(\mathbf{x}(w), \mathbf{y}(i-1), \hat{\mathbf{y}}_1(z|s_{i-1}), \mathbf{x}_1(s_{i-1})) \in A_\epsilon^{*(n)}$ for some $z \in S_{s_i}$. If such a unique w exists, then it is the decoded $\hat{w}_{i-1}$, otherwise the receiver declares an error.

The rate expression resulting from this decoding procedure is given by the following proposition.

Proposition 2: For the general relay channel, any rate R satisfying

$$R \leq I(X;Y|X_1) + \min\{I(X;\hat{Y}_1|X_1,Y), I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)\}$$

subject to

$$I(X_1;Y) \geq I(\hat{Y}_1;Y_1|X,X_1,Y) = I(\hat{Y}_1;Y_1|X_1,Y) - I(X;\hat{Y}_1|X_1,Y)$$

for some joint distribution $p(x,x_1,y,y_1,\hat{y}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)$, is achievable.

Proof: The details of the proof can be found in [26] and [36, Appendix A]. □

It can be shown using TS (see details in Appendix A) that the joint-decoding rate can be obtained as a special case of [2, Th. 6]. Moreover, for the same distribution chain, the rate of Proposition 1 is always at least as high as the joint-decoding rate of Proposition 2.

III. The Gaussian Relay Channel With Coded Modulation

In this section, we investigate the application of EAF with TS to the Gaussian relay channel. For this channel, the common practice is to use Gaussian codebooks and Gaussian quantization (GQ) at the relay [8], [18], [19]. The rate in Gaussian scenarios where coded modulation is applied is usually analyzed by applying DAF at the relay (see, for example, [23] and [22]). In this section, we show that when considering coded modulation, one should select the relay strategy according to the channel conditions: Gaussian selection seems a good choice when the SNR at the relay is low and DAF appears to be superior when the relay enjoys high SNR conditions. However, for intermediate SNR, there is much room for optimizing the estimation mapping at the relay. This is shown via numerical analysis with supporting analytical arguments.

In the following, we first recall the Gaussian relay channel with a Gaussian codebook, and then we consider the Gaussian relay channel under BPSK modulation constraint. Because we focus on the mapping at the relay, we consider here the Gaussian relay channel with an orthogonal relay-destination link of finite capacity C, also considered in [11]. This scenario is depicted in Fig. 2. In the context of the general relay channel with an orthogonal relay-destination link ,we note two papers. The first is the work of [37], which considered the relay channel with an orthogonal relay-destination link and with conditionally independent Y and Y_1, where furthermore, Y is a stochastically degraded version of Y_1. Another work we note is [38], which considers the orthogonal relay channel in which Y_1 is a deterministic function of (X,Y).

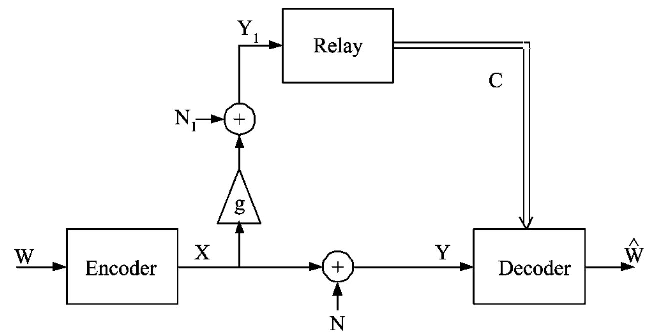

Fig. 2. Gaussian relay channel with a finite-capacity noiseless relay link between the relay and the destination.

In the Gaussian relay channel with an orthogonal relay-destination link, $Y_1 = g \cdot X + N_1$ is the channel output at the relay and $Y = X + N$ is the channel output at the receiver, which decodes the message based on Y^n and the information received from the relay through the relay-destination link. Here $g \in \mathbb{R}$, $N \sim \mathcal{N}(0,\sigma^2)$, and $N_1 \sim \mathcal{N}(0,\sigma_1^2)$, independent of N. Let $\mathcal{W} = \{1,2,\ldots,2^{nR}\}$ denote the source message set, and let the source have an average power constraint P

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \leq P \qquad \forall\, w \in \mathcal{W}.$$

The relay compresses its channel output Y_1^n into $\hat{Y}_1^n$ using the EAF scheme. The relay then sends information through the finite-capacity noiseless link to the destination to facilitate decoding of $\hat{Y}_1$ at the destination. For this scenario, the expressions of [2, Th. 6] specialize to

$$R \leq I(X;Y,\hat{Y}_1) \tag{14a}$$

$$\text{subject to } C \geq I(\hat{Y}_1;Y_1|Y) \tag{14b}$$

with the Markov chain $X,Y - Y_1 - \hat{Y}_1$.

We also consider in this section the DAF method whose achievable rate is given by (see [2, Th. 1])

$$R_{\text{DAF}} = \min\{I(X;Y_1), I(X;Y)+C\}$$

and the upper bound of [2, Th. 3]

$$R_{\text{upper}} = \min\{I(X;Y)+C, I(X;Y,Y_1)\}.$$

We note that although these expressions were originally derived for the finite and discrete alphabet case, following the method of Wyner in [39], they hold also for arbitrary sources, and in particular, for the Gaussian and the Gaussian-mixture cases.

A. The Gaussian Relay Channel With Gaussian Codebooks

When $X \sim \mathcal{N}(0,P)$, independent identically distributed (i.i.d.), then the channel outputs at the relay and at the receiver are jointly normal RVs

$$\begin{pmatrix} y \\ y_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P+\sigma^2 & gP \\ gP & g^2P+\sigma_1^2 \end{pmatrix} \right).$$

The compression is achieved by adding to Y_1 a zero mean independent Gaussian RV N_Q

$$\hat{Y}_1 = Y_1 + N_Q, \quad N_Q \sim \mathcal{N}\left(0, \sigma_Q^2\right). \tag{15}$$

We refer to the assignment (15) as Gaussian-quantization estimate-and-forward (GQ-EAF). Evaluating the expressions (14a) and (14b) with assignment (15) results in (see, also, [11] and [41])

$$I(X;Y,\hat{Y}_1) = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2} + \frac{g^2 P}{\sigma_1^2 + \sigma_Q^2}\right) \tag{16a}$$

$$I(Y_1;\hat{Y}_1|Y) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_1^2(\sigma^2+P) + g^2\sigma^2 P}{\sigma_Q^2(P+\sigma^2)}\right). \tag{16b}$$

The feasibility condition (14b), combined with (16b), yields

$$\sigma_Q^2 \geq \frac{\sigma_1^2(\sigma^2+P) + g^2\sigma^2 P}{(2^{2C}-1)(P+\sigma^2)}$$

and because maximizing the rate (16a) requires minimizing σ_Q^2, the resulting GQ-EAF rate expression is

$$R \leq \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2} + \frac{g^2 P}{\sigma_1^2 + \frac{\sigma_1^2(\sigma^2+P)+g^2\sigma^2 P}{(2^{2C}-1)(P+\sigma^2)}}\right).$$

Now, when using GQ at the relay, we see that TS does not help: the minimum σ_Q^2 is required in order to maximize the rate (16a). This minimum is achieved only when the entire capacity of the relay-destination link is dedicated to the transmission of the (minimally) quantized Y_1. However, when we consider the Gaussian relay channel with coded modulation, the situation is quite different, as we show in the remainder of this section.

B. The Gaussian Relay Channel With Coded Modulation

Consider the Gaussian relay channel where X is an equiprobable BPSK signal of amplitude $\sqrt{P}$

$$\Pr(X=\sqrt{P}) = \Pr(X=-\sqrt{P}) = \frac{1}{2}. \tag{17}$$

In this case, the received symbols (Y, Y_1) are no longer jointly Gaussian, but follow a Gaussian-mixture distribution

$$\begin{aligned} f(y,y_1) &= \Pr(X=\sqrt{P})f(y,y_1|x=\sqrt{P}) \\ &\quad + \Pr(X=-\sqrt{P})f(y,y_1|x=-\sqrt{P}) \\ &= \frac{1}{2}\Big(G_y(\sqrt{P},\sigma^2)G_{y_1}(g\sqrt{P},\sigma_1^2) \\ &\qquad + G_y(-\sqrt{P},\sigma^2)G_{y_1}(-g\sqrt{P},\sigma_1^2)\Big) \end{aligned}$$

where

$$G_x(a,b) \triangleq \frac{1}{\sqrt{2\pi b}}e^{-\frac{(x-a)^2}{2b}}. \tag{18}$$

Contrary to the Gaussian codebook case, where it is hard to identify a mapping $p(\hat{y}_1|y_1)$ that will be superior to the GQ (if indeed such a mapping exists), in this case, it is natural to consider ternary mappings for $\hat{Y}_1$. We can predict that such mappings will perform well at high SNR on the source–relay link, when the probability of error for symbol-by-symbol detection at the relay is small, with a much smaller complexity than GQ. We start by considering two types of HD mappings.

1) The first mapping is HD-EAF. The relay first makes a hard decision about every received Y_1 symbol, determining whether it is positive or negative, and then randomly decides whether to transmit this decision or transmit an erasure symbol E instead. The probability of transmitting an erasure $1 - P_{\text{no erase}}$ is used to adjust the conference rate such that the feasibility constraint is satisfied. Therefore, the conditional distribution $p(\hat{y}_1|y_1)$ is given by

$$p(\hat{y}_1|y_1>0) = \begin{cases} P_{\text{no erase}}, & \hat{y}_1 = 1 \\ 1-P_{\text{no erase}}, & \hat{y}_1 = E \end{cases} \tag{19a}$$

$$p(\hat{y}_1|y_1\leq 0) = \begin{cases} P_{\text{no erase}}, & \hat{y}_1 = -1 \\ 1-P_{\text{no erase}}, & \hat{y}_1 = E. \end{cases} \tag{19b}$$

This choice is motivated by the TS method considered in Section II: after making an HD on the received symbol's sign, positive or negative, the relay applies TS to that decision so that the rate required to transmit the resulting random variable to the destination is less than C. This facilitates transmission to the destination through the conference link. Because the entropy of the sign decision is 1, then when $C \geq 1$, we can transmit the sign decisions directly without using an erasure. Therefore, we expect that for values of C in the range $C > 1$, this mapping will not exceed the rate obtained for $C = 1$. The focus is, therefore, on values of C that are less than 1. The expressions for this assignment are given in part A of Appendix B.

2) The second method is deterministic HD. In this approach, we select a threshold T such that the range of Y_1 is partitioned into three regions: $y_1 < -T$, $-T \leq y_1 \leq T$, and $y_1 > T$. Then, according to the value of each received Y_1 symbol, the corresponding $\hat{Y}_1$ is deterministically selected

$$\hat{Y}_1 = \begin{cases} 1, & Y_1 > T \\ E, & -T \leq Y_1 \leq T \\ -1, & Y_1 < -T. \end{cases} \tag{20}$$

The threshold T is selected such that the achievable rate is maximized subject to satisfying the feasibility constraint. We refer to this method as deterministic hard decision (DHD). Therefore, this is another type of TS in which the erasure probability is determined by the fraction of the time the relay input is between $-T$ to T. The expressions for evaluating the rate of the DHD assignment are given in part B of Appendix B.

We expect the DHD method to be better than HD-EAF at high source–relay SNR because in HD-EAF, erasure is selected without any regard to the quality of the sign decision—both high-quality sign decisions and low-quality sign decisions are erased with the same probability. In contrast, in DHD, the erased region is the region where the decisions have low quality in the first place and all high-quality decisions are sent. However, at low source–relay SNR and small capacity of the relay-destination link, HD-EAF may perform better than DHD because the erased region (i.e., the region between $-T$ to $+T$) for the DHD mapping has to be very large in order to facilitate transmission of the estimate through the relay-destination link, while HD-EAF may require less compression of the HD output. This

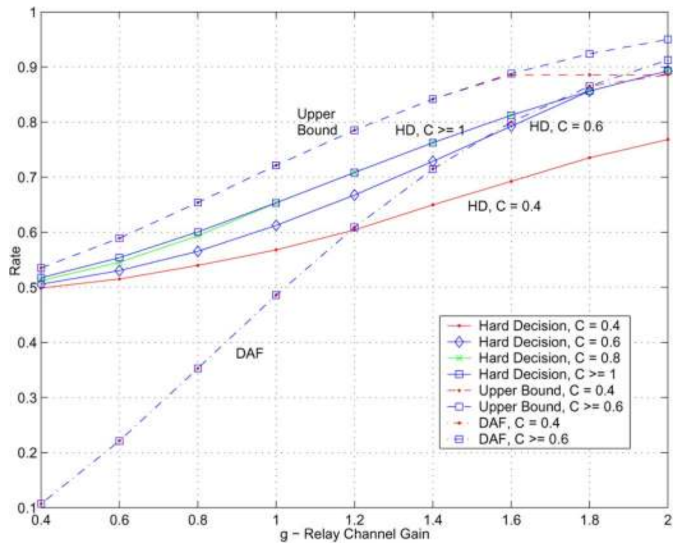

Fig. 3. Information rate with BPSK for HD-EAF mapping at the relay versus source–relay channel gain g, for different values of C.

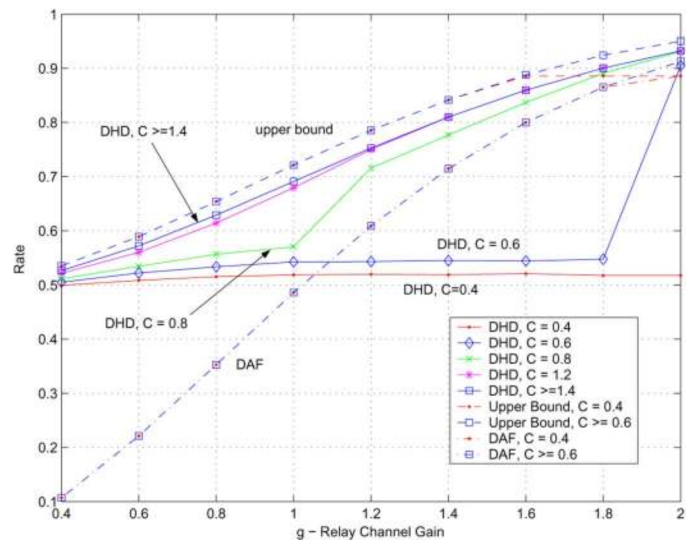

Fig. 4. Information rate with BPSK for DHD mapping at the relay versus source-relay channel gain g, for different values of C.

is because the erasure symbol in DHD carries information while in HD-EAF it does not. Therefore, DHD requires more bandwidth for transmission of this information to the destination.

We note here two related papers that consider relay transmission based on symbol-by-symbol decisions. The first paper is [21], which compares the performance obtained by soft and hard symbol-by-symbol decisions at the relay with AAF. The second work is [40] in which the bit error rate (BER) is compared for DAF and AAF where the focus is on the detector structure at the destination receiver.

We now examine the performance of each technique using numerical evaluation. First, we examine the achievable rates with HD-EAF. The expressions are evaluated for $\sigma_1^2 = \sigma^2 = 1$ and $P = 1$. For every pair of values (g, C) considered, the $P_{\text{no erase}}$ that maximizes the rate was selected. Fig. 3 depicts the information rate versus g for $0.4 \leq C \leq 2$, together with the upper bound and the DAF rate. As can be observed from Fig. 3, the information rate of HD-EAF increases with C until $C = 1$ and then remains constant (solid line with square markers). It is also seen that for small values of g, HD-EAF is better than DAF. This region of g increases with C, and for $C \geq 1$ the crossover value of g is approximately 1.71.

Next, examine DHD. As can be seen from Fig. 4, for small values of C, DAF exceeds the information rate of DHD for values of g greater than 1, but for $C \geq 0.8$, DHD is superior to DAF, and in fact DAF approaches DHD from below. Another phenomenon obvious from the figure (examine $C = 0.8$, for example) is the existence of a threshold. For low values of C there is some g at which the DHD rate exhibits a jump. This can be explained by looking at Fig. 5, which depicts the values of $I(X; Y, \hat{Y}_1)$ and $I(\hat{Y}_1; Y_1|Y)$ versus the threshold T: the bold–solid graph of $I(\hat{Y}_1; Y_1|Y)$ can intersect the horizontal bold–dashed line representing C at two values of T at the most. We also note that for small T the value of $I(X; Y, \hat{Y}_1)$ is generally larger than for large T. Now, the jump can be explained as follows. As shown in part B.1 of Appendix B, for small T and g, $I(\hat{Y}_1; Y_1|Y)$ is bounded from below. If this bound value is greater than C, then the intersection will occur only at a large value of T, hence the low information rate. When g increases, the value of $I(\hat{Y}_1; Y_1|Y)$ for small T decreases accordingly, until at some g it intersects C for a small T as well as for a large T, as indicated by the arrow in the right-hand part of Fig. 5. This allows us to obtain the rates in the region of small T, which are, in general, higher than the rates for large T and this is the source of the jump in the information rate.

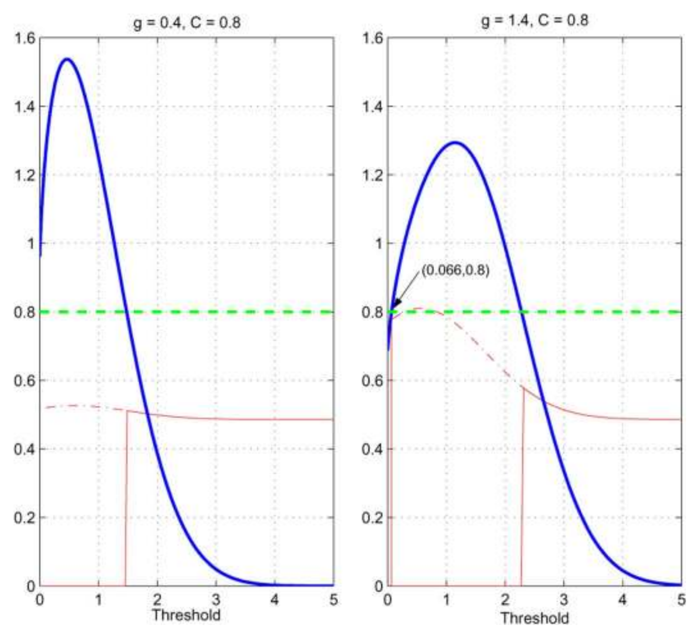

Fig. 5. $I(\hat{Y}_1; Y_1|Y)$ and $I(X; Y, \hat{Y}_1)$ versus threshold T for (left) $(g, C) = (0.4, 0.8)$ and (right) $(g, C) = (1.4, 0.8)$. The bold–solid line represents $I(\hat{Y}_1, Y_1|Y)$, the horizontal bold–dashed line represents $C = 0.8$, $I(X; Y, \hat{Y}_1)$ is represented by the dash–dot line, and the resulting information rate is depicted with the solid line.

C. Time-Sharing Deterministic Hard Decision (TS-DHD)

It is clearly evident from the above numerical evaluation that none of the two mappings, HD-EAF and DHD, is universally

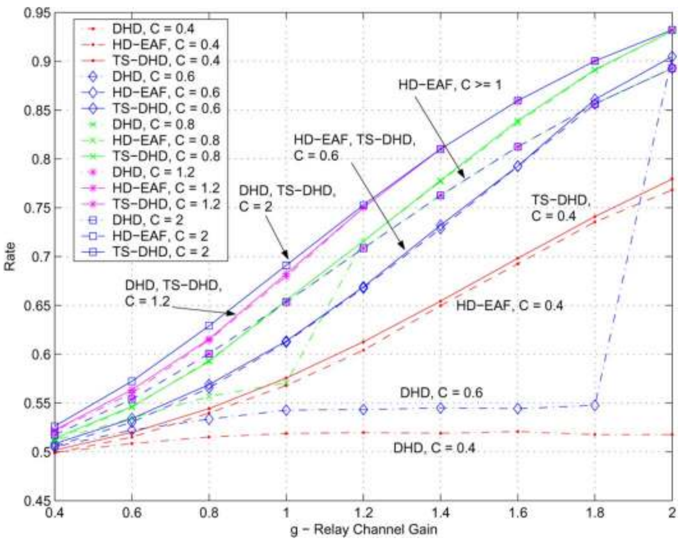

Fig. 6. Information rate with BPSK, for HD-EAF, DHD, and TS-DHD mappings at the relay versus relay channel gain g, for different values of C.

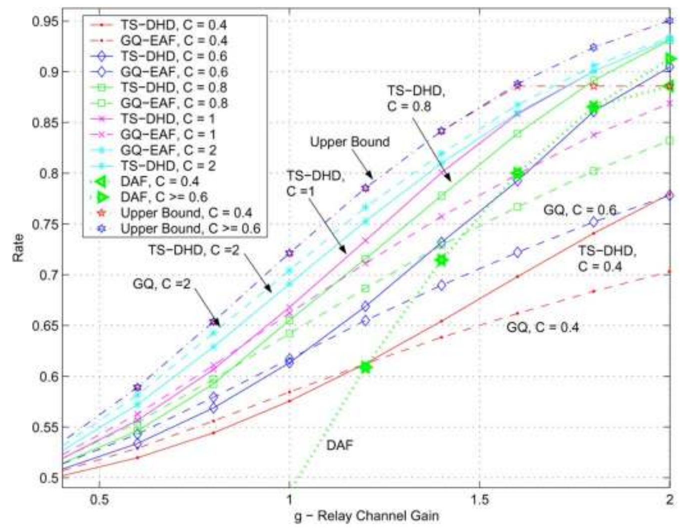

Fig. 7. Information rate with BPSK, for DAF, TS-DHD, and GQ-EAF at the relay versus relay channel gain g, for different values of C.

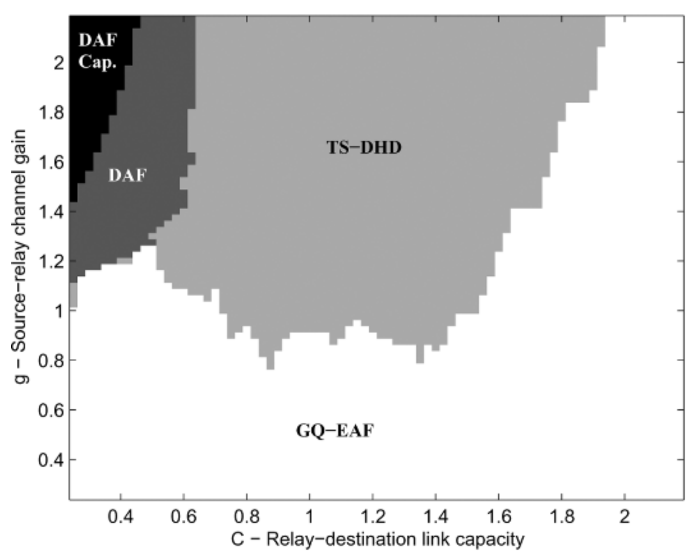

Fig. 8. Best relaying strategy (out of DAF, TS-DHD, and GQ-EAF) for the Gaussian relay channel with BPSK modulation.

better than the other: when g is small and C is less than 1, then HD-EAF performs better than DHD, because the erased region is too large, and when g increases, DHD performs better than HD-EAF because it erases only the low-quality information. It is, therefore, natural to consider a third mapping, which combines both aspects of binary mapping at the relay, namely, deterministically erasing low-quality information and then randomly gating (i.e., TS) the resulting discrete variable (in the region $|Y_1| > T$) in order to facilitate its transmission over the relay-destination link. This hybrid mapping is given in the following assignment:

$$p(\hat{y}_1|y_1 > T) = \begin{cases} P_{\text{no erase}}, & \hat{y}_1 = 1 \\ 1 - P_{\text{no erase}}, & \hat{y}_1 = E \end{cases} \tag{21a}$$

$$p(\hat{y}_1 = E|\,|y_1| \le T) = 1 \tag{21b}$$

$$p(\hat{y}_1|y_1 < -T) = \begin{cases} P_{\text{no erase}}, & \hat{y}_1 = -1 \\ 1 - P_{\text{no erase}}, & \hat{y}_1 = E. \end{cases} \tag{21c}$$

In this mapping, the region $|y_1| \le T$ is always erased, and the complement region is erased with probability $1 - P_{\text{no erase}}$. Of course, now both T and $P_{\text{no erase}}$ have to be optimized. The expressions for TS-DHD can be found in part C of Appendix B. Fig. 6 compares the performance of DHD, HD-EAF, and TS-DHD. As can be seen, the hybrid method enjoys the benefits of both types of mappings and is the superior method.

Next, Fig. 7 compares the performance of TS-DHD, GQ-EAF (see part D of Appendix B), and DAF. As can be seen from the figure, GQ is not always the optimal choice. For $C = 0.6$ (the lines with diamond-shaped markers), we can see that GQ-EAF is the best method for $g < 1.05$; for $1.05 < g < 1.55$, TS-DHD is the best method; and for $g > 1.55$, DAF achieves the highest rate. For $C = 1$ (x-shaped markers), TS-DHD is superior to both GQ-EAF and DAF for $g > 0.9$; and for $C = 2$, GQ-EAF is the superior method for all $g \le 2$. We conclude that for the practical Gaussian relay scenario, where the modulation constraint is taken into account, in some situations the three-level quantization is better than GQ.

Last, Fig. 8 depicts the regions in the $g - C$ plane in which each of the methods considered here is superior, in a similar manner to [11, Fig. 2].[3] As can be observed from the figure, in the noisy region of small g and also in the region of very large C (high C/g), GQ-EAF is superior, and in the strong relay region of medium-to-high g and medium-to-high C (medium C/g), TS-DHD is the superior method. DAF is superior for small C and high g (low C/g). We note that the region where DAF achieves capacity is obtained by numerically evaluating the upper and lower bounds on the rate. In some sense, the TS-DHD method is a hybrid method between the DAF, which makes an HD on the entire block, and GQ-EAF, which (can be thought of as if it) makes a soft decision every symbol (although it actually operates with blocks of n symbols), therefore it is superior in the transition region between the region where DAF is distinctly better, and the region where GQ-EAF is distinctly superior.

[3]The block shapes are due to the step size of 0.025 in the values of g and C at which the rates were evaluated.

D. When the SNR on the Source-Destination Link Approaches 0 $\left(\sigma^2 \rightarrow \infty\right)$

In this subsection, we analyze the relaying strategies discussed in this section as the SNR on the source-destination link $X-Y$ approaches zero. Because TS-DHD is a hybrid method combining both DHD and HD-EAF, we analyze the behavior of the components rather than the hybrid, to gain more insight. This analysis is particularly useful when trying to numerically evaluate the rates, because as the source-destination link SNR goes to zero, the computer's numerical accuracy does not allow to numerically evaluate the rates using the general expressions. This situation corresponds to two-hop relaying.

First, we note that when the SNR of the direct link $X-Y$ approaches 0, we have that $I(X;Y) \rightarrow 0$ as well. To see this, we write

$$I(X;Y) = h(Y) - h(X+N|X) = h(Y) - h(N)$$

with $h(Y) = -\int_{-\infty}^{\infty} f(y)\log_2(f(y))dy$, and from (B.3)

$$\begin{aligned} f(Y) &= \frac{1}{2}\left(G_y(\sqrt{P},\sigma^2) + G_y(-\sqrt{P},\sigma^2)\right) \\ &= \frac{1}{2}\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\sqrt{P})^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y+\sqrt{P})^2}{2\sigma^2}}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\left(\frac{1}{2}e^{\frac{y\sqrt{P}}{\sigma^2}} + \frac{1}{2}e^{-\frac{y\sqrt{P}}{\sigma^2}}\right)e^{-\frac{P}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right)e^{-\frac{P}{2\sigma^2}} \\ &\overset{\sigma^2\rightarrow\infty}{\approx} \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}} \\ &\triangleq G_y(0,\sigma^2) \end{aligned}$$

where the approximation is in the sense that, for small $|y|$, we have $\cosh(|y|) \approx 1$, and for large $|y|$, $e^{-\frac{y^2}{2\sigma^2}}$ drives the entire expression to zero as $e^{-\frac{y^2}{2\sigma^2}}$, for $\sigma^2 \rightarrow \infty$. This approximation reflects the intuitive notion that as the variance increases to infinity, the two-component, symmetric Gaussian mixture resembles more and more a zero-mean Gaussian RV with the same variance. Therefore, for low SNR, the output at the destination is very close to a zero-mean normal RV with variance σ^2, and $h(Y) \approx h(N)$,[4] hence

$$I(X;Y) \xrightarrow{\sigma^2\rightarrow\infty} 0.$$

Note that the upper bound and the DAF rate in this case are both equal to

$$R_{\text{DAF}} = R_{\text{upper}} = \min\left\{C, I(X;Y_1)\right\}.$$

Now, let us evaluate the rate for HD-EAF as the SNR goes to zero. In part E of Appendix B, we show that the rate for HD-EAF as the SNR on the source-destination link goes to zero becomes [see (B.10)]

$$R_{\text{HD-EAF}} \leq \min\left\{C, 1\right\}\left(1 - H(P_1, 1-P_1)\right) \qquad (22)$$

[4]For $\sigma = 20$, we have that $\int_{-\infty}^{\infty}|f_Y(y) - G_y(0,\sigma^2)|dy < 0.001$, for $\sigma = 55, h(Y) - h(N) \approx 0.001$, and for $\sigma = 200, h(Y) - h(N) < 0.0001$.

where $H(\mathbf{p})$ is the discrete entropy for the specified pdf $\mathbf{p}$, and $P_1 = \Pr(Y_1 > 0|X = \sqrt{P})$.

For GQ-EAF, we first approximate $f(Y,\hat{Y}_1)$ at low SNR beginning with (B.9)

$$\begin{aligned} &f_{Y,\hat{Y}_1}(y,\hat{y}_1) \\ &= \frac{1}{2}\Big(G_y(\sqrt{P},\sigma^2)G_{\hat{y}_1}(g\sqrt{P},\sigma_1^2+\sigma_Q^2) \\ &\qquad + G_y(-\sqrt{P},\sigma^2)G_{\hat{y}_1}(-g\sqrt{P},\sigma_1^2+\sigma_Q^2)\Big) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\Big(\frac{1}{2}G_{\hat{y}_1}(g\sqrt{P},\sigma_1^2+\sigma_Q^2)e^{\frac{y\sqrt{P}}{\sigma^2}} \\ &\qquad + \frac{1}{2}G_{\hat{y}_1}(-g\sqrt{P},\sigma_1^2+\sigma_Q^2)e^{\frac{-y\sqrt{P}}{\sigma^2}}\Big)e^{-\frac{P}{2\sigma^2}} \\ &\approx G_y(0,\sigma^2)f_{\hat{Y}_1}(\hat{y}_1) \end{aligned}$$

because the behavior of this expression versus y is largely determined by $G_y(0,\sigma^2)$, and $e^{\pm\frac{y\sqrt{P}}{\sigma^2}}$ has only a negligible effect. Again, we see that as the direct link SNR approaches 0, Y and $\hat{Y}_1$ become independent. Now, the rate is given by

$$\begin{aligned} R &\leq I(X;Y,\hat{Y}_1) \\ &\overset{(a)}{\approx} h(\hat{Y}_1) - h(N_1+N_Q) \end{aligned} \qquad (23)$$

where (a) is derived in part F of Appendix B. The feasibility condition becomes

$$\begin{aligned} C &\geq I(\hat{Y}_1;Y_1|Y) \\ &= h(\hat{Y}_1|Y) - h(\hat{Y}_1|Y,Y_1) \\ &\approx h(\hat{Y}_1) - h(N_Q) \end{aligned} \qquad (24)$$

with

$$f_{\hat{Y}_1}(\hat{y}_1) = \frac{1}{2}\left[G_{\hat{y}_1}(g\sqrt{P},\sigma_1^2+\sigma_Q^2) + G_{\hat{y}_1}(-g\sqrt{P},\sigma_1^2+\sigma_Q^2)\right].$$

Finally, for DHD, as $\sigma^2 \rightarrow \infty$, we have

$$\begin{aligned} I(X;\hat{Y}_1,Y) &= I(X;Y) + I(X;\hat{Y}_1|Y) \\ &\approx I(X;\hat{Y}_1|Y) \\ &= H(\hat{Y}_1|Y) - H(\hat{Y}_1|Y,X) \\ &\overset{(a)}{\approx} H(\hat{Y}_1) - H(\hat{Y}_1|X) \\ &= I(X;\hat{Y}_1) \end{aligned}$$

where (a) follows from the independence of Y and Y_1 as $\sigma^2 \rightarrow \infty$ and the fact that $\hat{Y}_1$ is a deterministic function of Y_1, combined with the fact that given X, Y_1 and Y are independent. The feasibility condition becomes

$$C \geq H(\hat{Y}_1|Y) \approx H(\hat{Y}_1).$$

Because $I(X;\hat{Y}_1)$ is not a monotone function of T, we have to optimize over T to find the actual rate.

As can be seen from the expression for HD-EAF, when the SNR on the source-destination link decreases, the capacity of the conference link acts as a scaling factor on the rate of the binary channel from the source to the relay. This is due to the

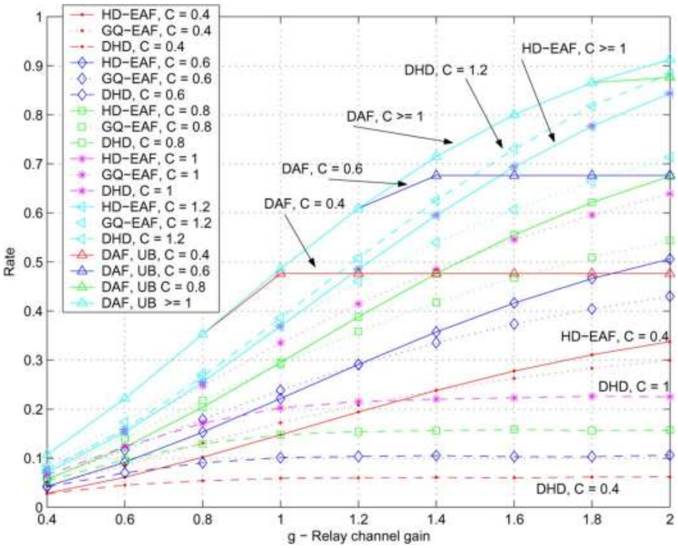

Fig. 9. Information rate with DAF, DHD, HD-EAF, and GQ-EAF versus source–relay channel gain g, for different values of C, at low SNR on the source-destination link.

TS. In Fig. 9, we plot the information rates for DHD, HD-EAF, GQ-EAF, and DAF (which coincides with the upper bound for asymptotically low SNR on the source-destination link). Comparing the three EAF strategies, we note that DHD, which at intermediate SNR on the source–relay link performs well for $C \geq 0.8$, has the worst performance at low SNR up to $C = 1.2$. At $C = 1.2$, DHD becomes the best scheme out of the three. This is due to the threshold effect discussed earlier. For $C < 1.2$ and high SNR on the source–relay channel, HD-EAF outperforms both DHD and GQ-EAF. For low SNR on the source–relay channel, GQ-EAF is again superior.

E. Discussion

We make the following observations.

- As noted at the beginning of this section, at low SNR on the source–relay link, GQ-EAF outperforms TS-DHD. To see why, consider the distribution of Y_1

$$f_{Y_1}(y_1) = G_{y_1}\left(0, \sigma_1^2\right) \cosh\left(\frac{g\sqrt{P} y_1}{\sigma_1^2}\right) e^{-\frac{g^2 P}{2\sigma_1^2}} \overset{g\to 0}{\approx} G_{y_1}\left(0, \sigma_1^2\right)$$

  where the approximation is because for large values of $Y_1, G_{y_1}\left(0, \sigma_1^2\right)$ determines the behavior of the expression. Therefore, as $g \to 0$, Y_1 approaches a zero-mean Gaussian RV: $Y_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2)$. As discussed in [28, Ch. 13.3.2], the rate-distortion function for quantizing a Gaussian RV is minimized by GQ (for squared error distortion). Therefore, it should be natural to guess that GQ will perform better at low SNR on the source–relay link.
  We also note that in [8, Sec. VII-B], EAF with a Gaussian auxiliary RV and Gaussian codebooks was evaluated for the general Gaussian relay channel. It was shown that at asymptotically high relay-destination SNR, this assignment of the codebooks and the auxiliary RV achieves capacity.
- At the other extreme, as $g \to \infty$, consider the DAF strategy: as $g \to \infty$, we have that $h(Y)_1$ is given by the equation shown at the bottom of the page, where the approximation is because as $g \to \infty$, the two Gaussian peaks in the Gaussian-mixture distribution are so far from one another that the effect of the overlap can be neglected. Therefore

$$\begin{aligned} I(X;Y_1) &= h(Y_1) - h(Y_1|X) \\ &\approx 1 + h(N_1) - h(N_1) \\ &= 1 \\ &= H(X). \end{aligned}$$

  Hence

$$\begin{aligned} R_{\text{DAF}} &= \min\{I(X;Y_1), I(X;Y) + C\} \\ &= \min\{1, I(X;Y) + C\} \end{aligned}$$

  which is the maximal rate. Therefore, as $g \to \infty$, DAF provides the optimal rate.
  This conclusion is in accordance with [8, Sec. VII-B], where it was shown that for the general Gaussian relay channel, DAF achieves capacity as the source–relay SNR goes to infinity. We note that for the general Gaussian relay channel, capacity is achieved with Gaussian codebooks. Here we showed that DAF maximizes the rate also for BPSK modulation.
- We can expect that at intermediate SNR, methods that combine elements of the "soft-decision per symbol" of GQ-EAF and the HD on the entire codeword of DAF, will be superior to both. As discussed earlier, TS-DHD is such a method. Furthermore, we believe that as the SNR decreases, increasing the cardinality of $\hat{Y}_1$ accordingly will improve the performance.
- We note that we did not make a comparison with the AAF scheme. The reason is that AAF generates an output variable X_1, which is a Gaussian RV. However, such an RV

$$\begin{aligned} h(Y_1) &= -\frac{1}{2}\int_{y_1=-\infty}^{\infty}\left[G_{y_1}(g\sqrt{P},\sigma_1^2) + G_{y_1}(-g\sqrt{P},\sigma_1^2)\right]\log_2\left(\frac{1}{2}\left[G_{y_1}(g\sqrt{P},\sigma_1^2) + G_{y_1}(-g\sqrt{P},\sigma_1^2)\right]\right)dy_1 \\ &\overset{g\to\infty}{\approx} 1 - \frac{1}{2}\int_{y_1=-\infty}^{\infty} G_{y_1}(g\sqrt{P},\sigma_1^2)\log_2 G_{y_1}(g\sqrt{P},\sigma_1^2)dy_1 - \frac{1}{2}\int_{y_1=-\infty}^{\infty} G_{y_1}(-g\sqrt{P},\sigma_1^2)\log_2 G_{y_1}(-g\sqrt{P},\sigma_1^2)dy_1 \\ &= 1 + h(N_1) \end{aligned}$$

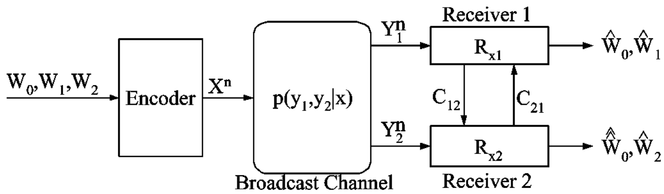

Fig. 10. BC with cooperating receivers. The encoder sends three messages, a common message W_0, a private message to Rx_1, W_1, and a private message to Rx_2, W_2. $\hat{W}_0$ and $\hat{\hat{W}}_0$ are the estimates of W_0 at Rx_1 and Rx_2, respectively.

cannot be transmitted through a finite-capacity link, therefore AAF is not applicable to this scenario.

IV. Multistep Cooperative Broadcast Application

In this section, we consider the cooperative broadcast scenario. In this scenario, one transmitter communicates with two receivers. In its most general form, the transmitter sends three independent messages: a common message intended for both receivers and two private messages, one for each receiver, where all three messages are encoded into a single-channel codeword X^n. Each receiver gets a noisy version of the codeword, Y_1^n at Rx_1 and Y_2^n at Rx_2. After reception, the receivers exchange messages in a K-cycle conference over noiseless conference links of finite capacities C_{12} and C_{21}. Each conference message is based on the channel output at each receiver and the conference messages previously received from the other receiver, in a similar manner to the conference defined by Willems in [30] for the cooperative multiple-access channel (MAC). After conferencing, each receiver decodes its message. This scenario is depicted in Fig. 10. This setup was studied in [13] for the single common message case over the independent BC (i.e., $p(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x}) = \prod_{i=1}^n p(y_{1,i}|x_i)p(y_{2,i}|x_i)$), and in [14], for the general setup with a single cycle of conferencing. The motivation for considering K cycles of conferencing comes from the fact that without knowledge of the other receiver's input (namely, with a single conferencing cycle), the conference messages necessarily contain also "noise," i.e., each receiver sends some information to the other receiver that does not help the other receiver in decoding its message (see also [14, Sec. IV-I.3]). Therefore, by using several cycles of conferencing, each receiver can enhance the other receiver's knowledge about its information, thus allowing that receiver to be more effective in helping the first receiver. Of course, this has to be done in a way such that the additional information rate gained for each unit of the conference bandwidth is more than that ratio for the previous conference steps, otherwise the additional cycles will not improve the overall rate compared with fewer cycles of conferencing.

A. Definitions

We use the standard definition for the discrete memoryless general BC given in [32]. We define a cooperative coding scheme as follows.

Definition 5: A (C_{12}, C_{21})-admissible K-cycle conference consists of the following elements.

1) K message sets from Rx_1 to Rx_2, denoted by $\mathcal{W}_{12}^{(1)}, \mathcal{W}_{12}^{(2)}, \ldots, \mathcal{W}_{12}^{(K)}$, and K message sets from Rx_2 to Rx_1, denoted by $\mathcal{W}_{21}^{(1)}, \mathcal{W}_{21}^{(2)}, \ldots, \mathcal{W}_{21}^{(K)}$. Message set $\mathcal{W}_{12}^{(k)}$ consists of $2^{nR_{12}^{(k)}}$ messages and message set $\mathcal{W}_{21}^{(k)}$ consists of $2^{nR_{21}^{(k)}}$ messages.
2) K mapping functions, one for each conference step from Rx_1 to Rx_2

$$h_{12}^{(k)} : \mathcal{Y}_1^n \times \mathcal{W}_{21}^{(1)} \times \mathcal{W}_{21}^{(2)} \times \cdots \times \mathcal{W}_{21}^{(k-1)} \mapsto \mathcal{W}_{12}^{(k)}$$

and K mapping functions, one for each conference step from Rx_2 to Rx_1

$$h_{21}^{(k)} : \mathcal{Y}_2^n \times \mathcal{W}_{12}^{(1)} \times \mathcal{W}_{12}^{(2)} \times \cdots \times \mathcal{W}_{12}^{(k)} \mapsto \mathcal{W}_{21}^{(k)}$$

for $k = 1, 2, \ldots, K$.

The conference rates satisfy

$$C_{12} = \sum_{k=1}^{K} R_{12}^{(k)}, \quad C_{21} = \sum_{k=1}^{K} R_{21}^{(k)}.$$

Definition 6: A $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, C_{12}, C_{21}, K)$ code for the general BC with a common message and two independent private messages consists of three sets of source messages $\mathcal{M}_0 = \{1, 2, \ldots, 2^{nR_0}\}$, $\mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\}$, and $\mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\}$, a mapping function at the transmitter

$$f : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \mapsto \mathcal{X}^n$$

a (C_{12}, C_{21})-admissible K-cycle conference, and two decoders

$$g_1 : \mathcal{W}_{21}^{(1)} \times \mathcal{W}_{21}^{(2)} \times \cdots \times \mathcal{W}_{21}^{(K)} \times \mathcal{Y}_1^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1$$
$$g_2 : \mathcal{W}_{12}^{(1)} \times \mathcal{W}_{12}^{(2)} \times \cdots \times \mathcal{W}_{12}^{(K)} \times \mathcal{Y}_2^n \mapsto \mathcal{M}_0 \times \mathcal{M}_2.$$

Definition 7: The average probability of error is defined as the average probability that at least one of the receivers does not decode its message pair correctly

$$P_e^{(n)} = \Pr\big(g_1(W_{21}^{(1)}, W_{21}^{(2)}, \ldots, W_{21}^{(K)}, Y_1^n) \neq (M_0, M_1) \text{ or } g_2(W_{12}^{(1)}, W_{12}^{(2)}, \ldots, W_{12}^{(K)}, Y_2^n) \neq (M_0, M_2)\big)$$

where we assume that each source message is selected uniformly and independently over its respective message set.

B. The Cooperative BC With Two Independent and One Common Message

We first present the general result for the cooperative broadcast scenario with a K-cycle conference. For $j = 1, 2$, denote with $\tilde{\mathbf{Y}}_j^{(k)} \triangleq (\hat{Y}_j^{(1)}, \hat{Y}_j^{(2)}, \ldots, \hat{Y}_j^{(k)})$, $\hat{\mathbf{Y}}_j \equiv \tilde{\mathbf{Y}}_j^{(K)}$, and $\tilde{\mathbf{y}}_j^{(k)} = (\hat{y}_j^{(1)}, \hat{y}_j^{(2)}, \ldots, \hat{y}_j^{(k)})$. Let R_1 and R_2 be the private rates to Rx_1 and Rx_2, respectively, and let R_0 denote the rate of the common information. Then, the following rate triplets are achievable.

Theorem 3: Consider the general BC $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ with cooperating receivers, having noiseless conference links of finite capacities C_{12} and C_{21} between them. Let the receivers hold a conference that consists of K cycles. Then, any rate triplet (R_0, R_1, R_2) satisfying

$$R_0 \le \min\{I(W;Y_1,\hat{\mathbf{Y}}_2), I(W;\hat{\mathbf{Y}}_1,Y_2)\} \quad (25a)$$
$$R_1 \le I(U;Y_1,\hat{\mathbf{Y}}_2|W) \quad (25b)$$
$$R_2 \le I(V;\hat{\mathbf{Y}}_1,Y_2|W) \quad (25c)$$
$$R_1+R_2 \le I(U;Y_1,\hat{\mathbf{Y}}_2|W)+I(V;\hat{\mathbf{Y}}_1,Y_2|W)-I(U;V|W) \quad (25d)$$

subject to

$$C_{12} \ge I(Y_1;\hat{\mathbf{Y}}_1,\hat{\mathbf{Y}}_2|Y_2) \quad (26a)$$
$$C_{21} \ge I(Y_2;\hat{\mathbf{Y}}_2,\hat{\mathbf{Y}}_1|Y_1) \quad (26b)$$

for some joint distribution

$$\begin{aligned} &p\Big(w,u,v,x,y_1,y_2,\tilde{\mathbf{y}}_1^{(K)},\tilde{\mathbf{y}}_2^{(K)}\Big) \\ &= p(w,u,v,x)p(y_1,y_2|x)p\left(\hat{y}_1^{(1)}|y_1\right)p\left(\hat{y}_2^{(1)}|y_2,\hat{y}_1^{(1)}\right)\cdots \\ &\times p\Big(\hat{y}_1^{(k)}|y_1,\tilde{\mathbf{y}}_1^{(k-1)},\tilde{\mathbf{y}}_2^{(k-1)}\Big)p\Big(\hat{y}_2^{(k)}|y_2,\tilde{\mathbf{y}}_1^{(k)},\tilde{\mathbf{y}}_2^{(k-1)}\Big)\cdots \\ &\times p\Big(\hat{y}_1^{(K)}|y_1,\tilde{\mathbf{y}}_1^{(K-1)},\tilde{\mathbf{y}}_2^{(K-1)}\Big)p\Big(\hat{y}_2^{(K)}|y_2,\tilde{\mathbf{y}}_1^{(K)},\tilde{\mathbf{y}}_2^{(K-1)}\Big) \end{aligned} \quad (27)$$

is achievable.

Proof:

1) Overview of the Strategy: The coding strategy is based on combining the broadcast code construction of [33], after incorporating the common message into the construction, with the K-cycle conference of [34]. The transmitter constructs a broadcast code to split the rate between the three message sets. This is done independently of the relaying scheme. Each receiver generates its conference messages according to the construction of [34]. After K cycles of conferencing, each receiver decodes its information based on its channel output and the conference messages received from the other receiver.

2) Code Construction at the Transmitter:

- Fix all the distributions in (27). Fix $\epsilon > 0$ and let $n > 1$. Let $\delta > 0$ be a positive number whose value is determined in the following steps. Let $R(W) = \min\{I(W;Y_1,\hat{\mathbf{Y}}_2), I(W;\hat{\mathbf{Y}}_1,Y_2)\}$. Let $S_{[W]\delta}^{(n)}$ denote the set of all $\mathbf{w} \in \mathcal{W}^n$ sequences such that $\mathbf{w} \in A_\delta^{*(n)}(W)$ and $A_\delta^{*(n)}(U,V|\mathbf{w})$ is nonempty, as defined in [27, Corollary 5.11]. From [27, Corollary 5.11], we have that $||S_{[W]\delta}^{(n)}|| \ge 2^{n(H(W)-\phi)}$, where $\phi \to 0$ as $\delta \to 0$ and $n \to \infty$.
- Pick $2^{n(R(W)-\epsilon)}$ sequences from $S_{[W]\delta}^{(n)}$ in a uniform and independent manner according to

$$\Pr(\mathbf{w}) = \begin{cases} \frac{1}{\left\|S_{[W]\delta}^{(n)}\right\|}, & \mathbf{w} \in S_{[W]\delta}^{(n)} \\ 0, & \text{otherwise.} \end{cases}$$

Label these sequences with $l \in \mathcal{M}_0 \triangleq \{1,2,\ldots,2^{n(R(W)-\epsilon)}\}$.

- For each sequence $\mathbf{w}(l)$, $l \in \mathcal{M}_0$, consider the set $A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$, $\delta' = \delta\max\{||\mathcal{U}||,||\mathcal{V}||\}$. Because the sequences $\mathbf{w} \in \mathcal{W}^n$ are selected such that $A_\delta^{*(n)}(U,V|\mathbf{w}(l))$ is nonempty and because $(\mathbf{u},\mathbf{v}) \in A_\delta^{*(n)}(U,V|\mathbf{w}(l))$ implies $\mathbf{u} \in A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$, then also $A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$ in nonempty, and by [27, Th. 5.9], $||A_{\delta'}^{*(n)}(U|\mathbf{w}(l))|| \ge 2^{n(H(U|W)-\psi)}$, $\psi \to 0$ as $\delta' \to 0$ and $n \to \infty$.
- For each $l \in \mathcal{M}_0$, pick $2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-\epsilon)}$ sequences in a uniform and independent manner from $A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$ according to

$$\Pr(\mathbf{u}|l) = \begin{cases} \frac{1}{\left\|A_{\delta'}^{*(n)}(U|\mathbf{w}(l))\right\|}, & \mathbf{u} \in A_{\delta'}^{*(n)}(U|\mathbf{w}(l)) \\ 0, & \text{otherwise.} \end{cases}$$

Label these sequences with

$$\mathbf{u}(i|l), \qquad i \in \mathcal{Z}_1 \triangleq \left\{1,2,\ldots,2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-\epsilon)}\right\}.$$

Similarly, pick $2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-\epsilon)}$ sequences in a uniform and independent manner from $A_{\delta'}^{*(n)}(V|\mathbf{w}(l))$ according to

$$\Pr(\mathbf{v}|l) = \begin{cases} \frac{1}{||A_{\delta'}^{*(n)}(V|\mathbf{w}(l))||}, & \mathbf{v} \in A_{\delta'}^{*(n)}(V|\mathbf{w}(l)) \\ 0, & \text{otherwise.} \end{cases}$$

Label these sequences with

$$\mathbf{v}(j|l), \qquad j \in \mathcal{Z}_2 \triangleq \left\{1,2,\ldots,2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-\epsilon)}\right\}.$$

δ and n are selected such that $||S_{[W]\delta}^{(n)}|| \ge 2^{n(R(W)-\epsilon)}$, and $\forall l \in \mathcal{M}_0$, we have that

$$||A_{\delta'}^{*(n)}(U|\mathbf{w}(l))|| \ge 2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-\epsilon)}$$
$$||A_{\delta'}^{*(n)}(V|\mathbf{w}(l))|| \ge 2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-\epsilon)}.$$

- Partition the set $\mathcal{Z}_1$ into 2^{nR_1} subsets B_{w_1}, $w_1 \in \mathcal{M}_1 = \{1,2,\ldots,2^{nR_1}\}$ and let

$$B_{w_1} = \Big[(w_1-1)2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-R_1-\epsilon)}+1,\; w_1 2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-R_1-\epsilon)}\Big].$$

Similarly, partition the set $\mathcal{Z}_2$ into 2^{nR_2} subsets C_{w_2}, $w_2 \in \mathcal{M}_2 = \{1,2,\ldots,2^{nR_2}\}$ and let

$$C_{w_2} = \Big[(w_2-1)2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-R_2-\epsilon)}+1,\; w_2 2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-R_2-\epsilon)}\Big].$$

- For each triplet (l,w_1,w_2), consider the set

$$\mathcal{D}(w_1,w_2|l) \triangleq \{(m_1,m_2): m_1 \in B_{w_1}, m_2 \in C_{w_2}, (\mathbf{u}(m_1|l),\mathbf{v}(m_2|l)) \in A_{\delta'}^{*(n)}(U,V|\mathbf{w}(l))\}.$$

By [33, Lemma, p. 121], we have that taking n large enough we can make $\Pr(||\mathcal{D}(w_1,w_2|l)||=0) \le \epsilon$ for any arbitrary $\epsilon > 0$, as long as

$$R_1 \le I(U;Y_1,\hat{\mathbf{Y}}_2|W) \quad (28a)$$
$$R_2 \le I(V;\hat{\mathbf{Y}}_1,Y_2|W) \quad (28b)$$
$$R_1+R_2 \le I(U;Y_1,\hat{\mathbf{Y}}_2|W)+I(V;\hat{\mathbf{Y}}_1,Y_2|W) - I(U;V|W). \quad (28c)$$

Note that the individual rate constraints are required to guarantee that the sets B_{w_1} and C_{w_2} are nonempty.

- For each $l \in \mathcal{M}_0$, we pick a unique pair of $(m_1(w_1,w_2,l), m_2(w_1,w_2,l)) \in \mathcal{D}(w_1,w_2|l)$, $(w_1,w_2) \in \mathcal{M}_1 \times \mathcal{M}_2$. The transmitter generates the codeword $\mathbf{x}(l,w_1,w_2)$ according to

$$p(\mathbf{x}(l,w_1,w_2)) = \prod_{i=1}^{n} p(x_i|u_i(m_1(w_1,w_2,l)), v_i(m_2(w_1,w_2,l)), w_i(l)).$$

When transmitting the triplet (l,w_1,w_2), the transmitter outputs $\mathbf{x}(l,w_1,w_2)$.

3) Codebook Construction at the Receivers:

- For the first conference step from Rx_1 to Rx_2, Rx_1 generates a codebook with $2^{nR_{12}'^{(1)}}$ codewords indexed by $z_{12}^{(1)} \in \mathcal{Z}_{12}^{(1)} = \{1,2,\ldots,2^{nR_{12}'^{(1)}}\}$ using the nth i.i.d. extension of the distribution $p(\hat{y}_1^{(1)})$: $p(\hat{\mathbf{y}}_1^{(1)}(z_{12}^{(1)})) = \prod_{i=1}^n p_{\hat{Y}_1}(\hat{y}_{1,i}^{(1)})$. Rx_1 uniformly and independently partitions the message set $\mathcal{Z}_{12}^{(1)}$ into $2^{nR_{12}^{(1)}}$ subsets indexed by $w_{12}^{(1)} \in \mathcal{W}_{12}^{(1)} = \{1,2,\ldots,2^{nR_{12}^{(1)}}\}$. Denote these subsets with $\mathcal{S}_{12,w_{12}^{(1)}}^{(1)}$.
- For the first conference step from Rx_2 to Rx_1, Rx_2 generates a codebook with $2^{nR_{21}'^{(1)}}$ codewords indexed by $z_{21}^{(1)} \in \mathcal{Z}_{21}^{(1)} = \{1,2,\ldots,2^{nR_{21}'^{(1)}}\}$ for each codeword $\hat{\mathbf{y}}_1^{(1)}(z_{12}^{(1)})$, $z_{12}^{(1)} \in \mathcal{Z}_{12}^{(1)}$, in an i.i.d. manner, i.e., the ith letter of the codeword $\hat{\mathbf{y}}_2^{(1)}(z_{21}^{(1)}|z_{12}^{(1)})$ is selected i.i.d. according to $p_{\hat{Y}_2|\hat{Y}_1}(\hat{y}_{2,i}^{(1)}|\hat{y}_{1,i}^{(1)}(z_{12}^{(1)}))$. Rx_2 uniformly and independently partitions the message set $\mathcal{Z}_{21}^{(1)}$ into $2^{nR_{21}^{(1)}}$ subsets indexed by $w_{21}^{(1)} \in \mathcal{W}_{21}^{(1)} = \{1,2,\ldots,2^{nR_{21}^{(1)}}\}$. Denote these subsets with $\mathcal{S}_{21,w_{21}^{(1)}}^{(1)}$.
- For the kth conference step from Rx_1 to Rx_2, Rx_1 considers each combination of $\tilde{\mathbf{z}}_{12}^{(k-1)}, \tilde{\mathbf{z}}_{21}^{(k-1)}$, $\tilde{\mathbf{z}}_{12}^{(k-1)} \triangleq (z_{12}^{(1)}, z_{12}^{(2)}, \ldots, z_{12}^{(k-1)})$, $\tilde{\mathbf{z}}_{21}^{(k-1)} \triangleq (z_{21}^{(1)}, z_{21}^{(2)}, \ldots, z_{21}^{(k-1)})$. For each combination, Rx_1 generates a codebook with $2^{nR_{12}'^{(k)}}$ messages indexed by $z_{12}^{(k)} \in \mathcal{Z}_{12}^{(k)} = \{1,2,\ldots,2^{nR_{12}'^{(k)}}\}$, according to the distribution $p(\hat{y}_1^{(k)}|\hat{y}_1^{(1)},\hat{y}_1^{(2)},\ldots,\hat{y}_1^{(k-1)},\hat{y}_2^{(1)},\hat{y}_2^{(2)},\ldots,\hat{y}_2^{(k-1)})$. Rx_1 uniformly and independently partitions the message set $\mathcal{Z}_{12}^{(k)}$ into $2^{nR_{12}^{(k)}}$ subsets indexed by $w_{12}^{(k)} \in \mathcal{W}_{12}^{(k)} = \{1,2,\ldots,2^{nR_{12}^{(k)}}\}$. Denote these subsets with $\mathcal{S}_{12,w_{12}^{(k)}}^{(k)}$.
- The codebook for the kth conference step from Rx_2 to Rx_1 is generated in a parallel manner for each combination of $z_{12}^{(1)}, z_{12}^{(2)}, \ldots, z_{12}^{(k)}, z_{21}^{(1)}, z_{21}^{(2)}, \ldots, z_{21}^{(k-1)}$.

4) Decoding and Encoding at Rx_1 *at the* k*th Conference Cycle* $(k \le K)$ *for Transmission Block* i*:* Rx_1 needs first to decode the message $z_{21}^{(k-1)}$ sent from Rx_2 at the $(k-1)$th cycle. To that end, Rx_1 uses $w_{21}^{(k-1)}$, the index received from Rx_2 at the $(k-1)$th conference step. In decoding $z_{21}^{(k-1)}$, we assume that all the previous $z_{21}^{(1)}, z_{21}^{(2)}, \ldots, z_{21}^{(k-2)}$ were correctly decoded at Rx_1. We denote the $\hat{\mathbf{y}}_2^{(j)}$ sequences corresponding to $z_{21}^{(1)}, z_{21}^{(2)}, \ldots, z_{21}^{(k-2)}$ by $\hat{\mathbf{y}}_2(1), \hat{\mathbf{y}}_2(2), \ldots, \hat{\mathbf{y}}_2(k-2)$, and similarly, define $\hat{\mathbf{y}}_1(1), \hat{\mathbf{y}}_1(2), \ldots, \hat{\mathbf{y}}_1(k-1)$.

- Rx_1 first generates the set $\mathcal{L}_1(k-1)$ defined by the equation shown at the bottom of the page.
- Rx_1 then looks for a unique $z_{21}^{(k-1)} \in \mathcal{Z}_{21}^{(k-1)}$ such that $z_{21}^{(k-1)} \in \mathcal{L}_1(k-1) \bigcap \mathcal{S}_{21,w_{21}^{(k-1)}}^{(k-1)}$. If there is none or there is more than one, an error is declared.
- From an argument similar to the derivation of [34, eq. (3.5)], the probability of error can be made arbitrarily small by taking n large enough as long as

$$R_{21}'^{(k-1)} < I(\hat{Y}_2^{(k-1)}; Y_1|\tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-2)}) + R_{21}^{(k-1)} - \epsilon.$$

Here, $k > 1$, because the first conference message from Rx_1 to Rx_2 (which is also the first conference message in the K-cycle conference) is generated based only on the channel output at Rx_1, $\mathbf{y}_1(i)$, because there is no message from Rx_2 to decode for this conference step.

In generating the kth conference message to Rx_2, it is assumed that all the previous $k-1$ messages from Rx_2 were decoded correctly.

- Rx_1 looks for a message $z_{12}^{(k)} \in \mathcal{Z}_{12}^{(k)}$ such that

$$\Big(\hat{\mathbf{y}}_1^{(k)}(z_{12}^{(k)}|\tilde{\mathbf{z}}_{12}^{(k-1)}, \tilde{\mathbf{z}}_{21}^{(k-1)}), \hat{\mathbf{y}}_1(1), \hat{\mathbf{y}}_1(2), \ldots, \hat{\mathbf{y}}_1(k-1), \hat{\mathbf{y}}_2(1), \hat{\mathbf{y}}_2(2), \ldots, \hat{\mathbf{y}}_2(k-1), \mathbf{y}_1(i)\Big) \in A_\epsilon^{*(n)}.$$

From the argument in [34, eq. (3.3)], the probability that such a sequence exists can be made arbitrarily close to 1 by taking n large enough as long as

$$R_{12}'^{(k)} > I(\hat{Y}_1^{(k)}; Y_1|\tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) + \epsilon.$$

- Rx_1 looks for the partition of $\mathcal{Z}_{12}^{(k)}$ into which $z_{12}^{(k)}$ belongs. Denote the index of this partition with $w_{12}^{(k)}$.
- Rx_1 transmits $w_{12}^{(k)}$ to Rx_2 through the conference link.

5) Decoding and Encoding at Rx_2 *at the* k*th Conference Step* $(k \le K)$ *for Transmission Block* i*:* Using similar arguments to Section IV-B4, we obtain the following rate constraints.

- Decoding $z_{12}^{(k)}$ at Rx_2 can be done with an arbitrarily small probability of error by taking n large enough as long as

$$R_{12}'^{(k)} < I\left(\hat{Y}_1^{(k)}; Y_2|\tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}\right) + R_{12}^{(k)} - \epsilon.$$

$$\mathcal{L}_1(k-1) = \Big\{z_{21}^{(k-1)} \in \mathcal{Z}_{21}^{(k-1)} : \Big(\hat{\mathbf{y}}_2^{(k-1)}(z_{21}^{(k-1)}|\tilde{\mathbf{z}}_{12}^{(k-1)}, \tilde{\mathbf{z}}_{21}^{(k-2)}), \hat{\mathbf{y}}_1(1), \hat{\mathbf{y}}_1(2), \ldots, \hat{\mathbf{y}}_1(k-1), \hat{\mathbf{y}}_2(1), \hat{\mathbf{y}}_2(2), \ldots, \hat{\mathbf{y}}_2(k-2), \mathbf{y}_1(i)\Big) \in A_\epsilon^{*(n)}\Big\}.$$

- Encoding $z_{21}^{(k)}$ can be done with an arbitrarily small probability of error by taking n large enough as long as

$$R_{21}'^{(k)} > I\left(\hat{Y}_2^{(k)}; Y_2 | \tilde{\mathbf{Y}}_1^{(k)}, \tilde{\mathbf{Y}}_2^{(k-1)}\right) + \epsilon.$$

6) Combining All Conference Rate Bounds: First, consider the bounds on $R_{12}'^{(k)}$, $k = 1, 2, \dots, K$

$$I\left(\hat{Y}_1^{(k)}; Y_1 | \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}\right) + \epsilon < R_{12}'^{(k)} < I\left(\hat{Y}_1^{(k)}; Y_2 | \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}\right) + R_{12}^{(k)} - \epsilon.$$

This can be satisfied only if

$$\begin{aligned} & I(\hat{Y}_1^{(k)}; Y_2 | \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) + R_{12}^{(k)} - \epsilon \\ & \quad > I(\hat{Y}_1^{(k)}; Y_1 | \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) + \epsilon \\ \Rightarrow R_{12}^{(k)} & > H(\hat{Y}_1^{(k)} | Y_2, \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) \\ & \quad - H(\hat{Y}_1^{(k)} | Y_1, \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) + 2\epsilon \\ & = I(\hat{Y}_1^{(k)}; Y_1 | Y_2, \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) + 2\epsilon. \end{aligned}$$

Hence

$$\begin{aligned} C_{12} &= \sum_{k=1}^{K} R_{12}^{(k)} \\ &\geq \sum_{k=1}^{K} \left(I(\hat{Y}_1^{(k)}; Y_1 | Y_2, \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) + 2\epsilon \right) \\ &= \sum_{k=1}^{K} \Big[I(\hat{Y}_1^{(k)}; Y_1 | Y_2, \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}) \\ &\qquad + I(\hat{Y}_2^{(k)}; Y_1 | Y_2, \tilde{\mathbf{Y}}_1^{(k)}, \tilde{\mathbf{Y}}_2^{(k-1)}) \Big] + 2K\epsilon \\ &= \sum_{k=1}^{K} I\left(\hat{Y}_1^{(k)}, \hat{Y}_2^{(k)}; Y_1 | Y_2, \tilde{\mathbf{Y}}_1^{(k-1)}, \tilde{\mathbf{Y}}_2^{(k-1)}\right) + 2K\epsilon \\ &= I(\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2; Y_1 | Y_2) + 2K\epsilon \end{aligned} \tag{29}$$

and, similarly

$$C_{21} \geq I(\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2; Y_2 | Y_1) + 2K\epsilon. \tag{30}$$

This provides the rate constraints on the conference auxiliary variables of (26a) and (26b).

7) Decoding at Rx_1*:* Rx_1 uses $\mathbf{y}_1(i)$ and $\hat{\mathbf{y}}_2^{(1)}, \hat{\mathbf{y}}_2^{(2)}, \dots, \hat{\mathbf{y}}_2^{(K)}$ received from Rx_2, to decode $(l_i, w_{1,i})$ as follows: Rx_1 looks for a unique message $l \in \mathcal{M}_0$ such

$$(\mathbf{w}(l), \mathbf{y}_1(i), \hat{\mathbf{y}}_2^{(1)}, \hat{\mathbf{y}}_2^{(2)}, \dots, \hat{\mathbf{y}}_2^{(K)}) \in A_\epsilon^{*(n)}.$$

From the point-to-point channel capacity theorem (see [33]), this can be done with an arbitrarily small probability of error by taking n large enough as long as

$$R_0 \leq I(W; Y_1, \hat{\mathbf{Y}}_2). \tag{31}$$

Denote the decoded message $\hat{l}_i$. Now Rx_1 decodes $w_{1,i}$ by looking for a unique $k \in \mathcal{Z}_1$ such that

$$(\mathbf{u}(k|\hat{l}_i), \mathbf{w}(\hat{l}_i), \mathbf{y}_1(i), \hat{\mathbf{y}}_2^{(1)}, \hat{\mathbf{y}}_2^{(2)}, \dots, \hat{\mathbf{y}}_2^{(K)}) \in A_\epsilon^{*(n)}.$$

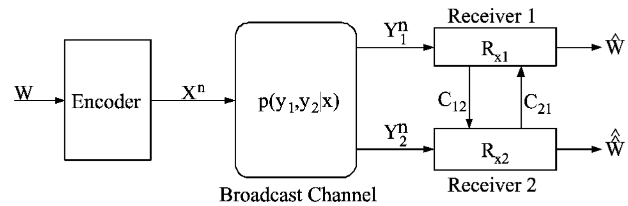

Fig. 11. BC with cooperating receivers, for the single common message case. $\hat{W}$ and $\hat{\hat{W}}$ are the estimates of W at Rx_1 and Rx_2, respectively.

If such unique k exists, then denote the decoded index with $\hat{k} = k$. Now Rx_1 looks for the partition of $\mathcal{Z}_1$ into which $\hat{k}$ belongs and sets $\hat{w}_{1,i}$ to be the index of that partition: $\hat{k} \in B_{\hat{w}_{1,i}}$. Similarly to the proof in [28, Ch 14.6.2], assuming successful decoding of l_i, the probability of error can be made arbitrarily small by taking n large enough as long as

$$\frac{1}{n} \log_2 \|\mathcal{Z}_1\| \leq I(U; Y_1, \hat{\mathbf{Y}}_2 | W)$$

which is satisfied by construction.

8) Decoding at Rx_2*:* Repeating similar steps for decoding at Rx_2, we get that decoding l_i can be done with an arbitrarily small probability of error by taking n large enough as long as

$$R_0 \leq I(W; \hat{\mathbf{Y}}_1, Y_2) \tag{32}$$

and assuming successful decoding of l_i, decoding $w_{2,i}$ with an arbitrarily small probability of error requires that

$$\frac{1}{n} \log_2 \|\mathcal{Z}_2\| \leq I(V; \hat{\mathbf{Y}}_1, Y_2 | W)$$

which again is satisfied by construction. Finally, collecting (28a)–(28c), (31) and (32) give the achievable rate constraints of Theorem 3, and (29) and (30) give the conference rate constraints of the theorem.

C. The Cooperative BC With a Single Common Message

In the single common message cooperative broadcast scenario, a single transmitter sends a message to two receivers encoded in a single channel codeword X^n. This scenario is depicted in Fig. 11. After conferencing, each receiver decodes the message. For this setup, we have the following upper bound.

Proposition 3 [31, Th. 6]: Consider the general BC $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ with cooperating receivers having noiseless conference links of finite capacities C_{12} and C_{21} between them. Then, for sending a common message to both receivers, any rate R must satisfy

$$R \leq \sup_{p_X(x)} \min \left\{ I(X; Y_1) + C_{21}, I(X; Y_2) + C_{12}, I(X; Y_1, Y_2) \right\}.$$

In [31], we also derived the following achievable rate for this scenario.

Proposition 4 [31, Th. 5]: Assume the BC setup of Proposition 3. Then, for sending a common message to both receivers, any rate R satisfying

$$R \leq \sup_{p_X(x)} \Big[\max\{R_{12}(p_X(x)), R_{21}(p_X(x))\} \Big]$$

$$R_{12}(p_X(x)) \triangleq \min\Big(I(X;Y_1) + C_{21}, I(X;Y_2) + \max\{0, m_{12} - H(Y_1|Y_2,X)\} \Big) \quad (33a)$$

$$R_{21}(p_X(x)) \triangleq \min\Big(I(X;Y_2) + C_{12}, I(X;Y_1) + \max\{0, m_{21} - H(Y_2|Y_1,X)\} \Big)$$

$$m_{ij} = \min(C_{ij}, H(Y_i|Y_j)) \quad (33b)$$

is achievable.

Note that this rate expression depends only on the parameters of the problem and is, therefore, computable. In Proposition 4, the achievable rate increases linearly with the cooperation capacity. The downside of this method is that it produces a rate increase over the noncooperative rate only for conference links capacities that exceed some minimum values.

Specializing the three independent messages result to the single common message case, we obtain the following achievable rate with a K-cycle conference for the general BC with a single common message.

Corollary 2: Consider the general BC with cooperating receivers, having noiseless conference links of finite capacities C_{12} and C_{21} between them. Let the receivers hold a conference that consists of K cycles. Then, any rate R satisfying

$$R = \max\{R_{12}, R_{21}\} \quad (34)$$

is achievable.

Here R_{12} is defined as follows:

$$R_{12} = \sup_{p_X(x), \alpha \in [0,1]} \min\{R_1, R_2\} \quad (35)$$

with

$$R_1 = I(X; Y_1, \tilde{\mathbf{Y}}_2^{(K-1)}) + \alpha C_{21} \quad (36a)$$

$$R_2 = I(X; Y_2, \tilde{\mathbf{Y}}_1^{(K)}) \quad (36b)$$

subject to

$$C_{12} \geq I(Y_1; \tilde{\mathbf{Y}}_1^{(K)}, \tilde{\mathbf{Y}}_2^{(K-1)} | Y_2) \quad (37a)$$

$$(1-\alpha) C_{21} \geq I(Y_2; \tilde{\mathbf{Y}}_1^{(K)}, \tilde{\mathbf{Y}}_2^{(K-1)} | Y_1) \quad (37b)$$

for the joint distribution

$$\begin{aligned} p(x, y_1, y_2, &\tilde{\mathbf{y}}_1^{(K)}, \tilde{\mathbf{y}}_2^{(K-1)}) \\ &= p(x)p(y_1,y_2|x)p(\hat{y}_1^{(1)}|y_1)p(\hat{y}_2^{(1)}|y_2,\hat{y}_1^{(1)}) \cdots \\ &\times p(\hat{y}_1^{(k)}|y_1, \tilde{\mathbf{y}}_1^{(k-1)}, \tilde{\mathbf{y}}_2^{(k-1)}) p(\hat{y}_2^{(k)}|y_2, \tilde{\mathbf{y}}_1^{(k)}, \tilde{\mathbf{y}}_2^{(k-1)}) \cdots \\ &\times p(\hat{y}_2^{(K-1)}|y_2, \tilde{\mathbf{y}}_1^{(K-1)}, \tilde{\mathbf{y}}_2^{(K-2)}) \\ &\times p(\hat{y}_1^{(K)}|y_1, \tilde{\mathbf{y}}_1^{(K-1)}, \tilde{\mathbf{y}}_2^{(K-1)}). \end{aligned}$$

R_{21} is defined in a parallel manner to R_{12}, with Rx_2 performing the first conference step, and the appropriate change in the probability chain.

The proof of Corollary 2 is provided in Appendix C.

We note that [13, Th. 2] presents a similar result for this scenario, under the constraint that the memoryless BC can be decomposed as $p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) = \prod_{i=1}^n p(y_{1,i}|x_i)p(y_{2,i}|x_i)$, and considering the sum rate of the conference. Here we show that the same achievable rate expressions hold for the general memoryless BC.

D. A Single-Cycle Conference With TS-EAF

Consider the case where only a single cycle of conferencing between the receivers is allowed. Specializing Corollary 2 to a single-cycle case, we obtain

$$R_1 = I(X;Y_1) + C_{21} \quad (38a)$$

$$R_2 = I(X;Y_2,\hat{Y}_1^{(1)}) \quad (38b)$$

$$C_{12} \geq I(Y_1;\hat{Y}_1^{(1)}|Y_2) \quad (38c)$$

and the TS-EAF assignment is

$$p(\hat{y}_1^{(1)}|y_1) = \begin{cases} q_1, & \hat{y}_1^{(1)} = y_1 \\ 1-q_1, & \hat{y}_1^{(1)} = E \notin \mathcal{Y}_1. \end{cases}$$

Applying the TS-EAF assignment to (38c) and (38b), we obtain

$$\begin{aligned} C_{12} &\geq I(Y_1;\hat{Y}_1^{(1)}|Y_2) \\ &= H(Y_1|Y_2) - H(Y_1|Y_2,\hat{Y}_1^{(1)}) \\ &= q_1 H(Y_1|Y_2) \\ R_2 &= I(X;Y_2,\hat{Y}_1^{(1)}) \\ &= I(X;Y_2) + H(X|Y_2) - H(X|Y_2,\hat{Y}_1^{(1)}) \\ &= I(X;Y_2) + q_1 I(X;Y_1|Y_2). \end{aligned}$$

Maximizing R_2 requires maximizing $q_1 \in [0,1]$. Therefore, setting $q_1 = \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^*$, we obtain $R_2 = I(X;Y_2) + \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^* I(X;Y_1|Y_2)$. Combining with R_1, we have that the rate when Rx_2 decodes first is given by

$$R_{12} = \min\left\{ I(X;Y_1) + C_{21}, I(X;Y_2) + \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^* I(X;Y_1|Y_2) \right\}$$

and by a symmetric argument, we can obtain R_{21}. We conclude that the rate for the single-cycle conference with TS-EAF is given by

$$\begin{aligned} R &= \max\left\{ \sup_{p(x)} R_{12}, \sup_{p(x)} R_{21} \right\} \\ R_{12} &= \min\left\{ I(X;Y_1) + C_{21}, I(X;Y_2) + \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^* I(X;Y_1|Y_2) \right\} \\ R_{21} &= \min\left\{ I(X;Y_1) + \left[\frac{C_{21}}{H(Y_2|Y_1)}\right]^* I(X;Y_2|Y_1), I(X;Y_2) + C_{12} \right\}. \end{aligned}$$

We note that this rate is always higher than the point-to-point rate and is also higher than the joint-decoding rate of Proposition 4 (whenever cooperation can provide a rate increase). However, as in Proposition 4, at least one receiver has to satisfy

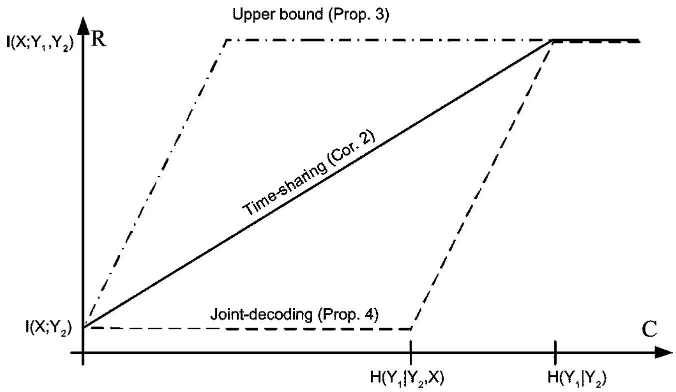

Fig. 12. Achievable rate R versus conference capacity C, for Proposition 3 (dashed–dotted), Proposition 4 (dashed), and Corollary 2 (solid), for the symmetric BC.

the Slepian–Wolf condition for the full cooperation rate to be achieved.

We demonstrate the results of Proposition 4 and Corollary 2 through a symmetric BC example: consider the symmetric BC where $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Y}$ and

$$p_{Y_1|Y_2,X}(a|b,x) = p_{Y_2|Y_1,X}(a|b,x)$$

for any $(a,b) \in \mathcal{Y} \times \mathcal{Y}$ and $x \in \mathcal{X}$. Let $C_{21} = C_{12} = C$. For this scenario, we have that $R_{12} = R_{21}$, in Corollary 2 and also $R_{12}(p_X(x)) = R_{21}(p_X(x))$ in Proposition 4. The resulting rate is depicted in Fig. 12 for a fixed probability $p(x)$. We can see that for this case, TS exceeds joint decoding for all values of C. Both methods meet the upper bound at $C = H(Y_1|Y_2)$. We note that this a corrected version of the figure in [35] (the figure in [35] is incorrect due to an error in the derivation).

We also note that using TS-EAF with more than two steps does not improve upon this result. For the three-step conference, for example, we achieve the following rate.

Corollary 3: For the general BC of Corollary 2, any rate R satisfying

$$R = \max\left\{ \sup_{p_X(x),\alpha\in[0,1]} \min\{R_1^{(212)}, R_2^{(212)}\}, \sup_{p_X(x),\alpha\in[0,1]} \min\{R_1^{(121)}, R_2^{(121)}\} \right\}$$

$$R_1^{(212)} = I(X;Y_1) + \left[\frac{\alpha C_{21}}{H(Y_2|Y_1)}\right]^* I(X;Y_2|Y_1) + (1-\alpha)C_{21} \tag{39a}$$

$$R_2^{(212)} = I(X;Y_2) + \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^* I(X;Y_1|Y_2) \tag{39b}$$

is achievable, where $R^{(iji)}$ indicates that the first step is performed by Rx_i, the second step by Rx_j and so on, and $R_1^{(121)}$ can be obtained from $R_2^{(212)}$ by switching "1" and "2," and similarly, $R_2^{(121)}$ can be obtained from $R_1^{(212)}$.

Examining the rate expression for $R_1^{(212)}$, we see that the maximum is achieved for $\alpha = 0$, namely, Rx_2 does not send any information in the first step and keeps all the conference bandwidth C_{21} for the last step. The reason is that using TS-EAF, every unit of cooperation capacity translates into additional rate of $\frac{1}{H(Y_2|Y_1)} I(X;Y_2|Y_1)$, up to maximum of $\frac{C_{21}}{H(Y_2|Y_1)} I(X;Y_2|Y_1)$ (assume $C_{21} < H(Y_2|Y_1)$). However, using DAF, every unit of conference capacity translates directly into a unit increase in the achievable rate. Therefore, DAF is more efficient than the TS-EAF step. Furthermore, in TS-EAF, using several cycles of conferencing does not allow to successively refine the conference information, because at every conference step, say of capacity C', the TS-EAF effectively delivers $n\frac{C'}{H(Y_2|Y_1)}$ output symbols from one receiver to the other. Once these symbols have been delivered, they do not participate in the following "successive refinements" as the conditional distribution becomes atomic. Therefore, there is no gain in performing several TS-EAF steps compared with a single TS-EAF step. Note that this may not hold when the mapping is not TS-EAF.

V. Conclusion

In this paper, we considered the EAF technique using TS on the auxiliary RVs. We first presented an alternative characterization of the classic EAF rate without a feasibility constraint. We also showed that incorporating joint decoding at the destination into the EAF scheme results in a special case of the classic EAF of [2, Th. 6]. Next, we showed that for the Gaussian relay channel with coded modulation, the rate achievable with the Gaussian auxiliary RV assignment can be exceeded by a TS-EAF implementing a "per-symbol HD," under certain channel conditions. Finally, we considered a third application of TS-EAF to the cooperative broadcast scenario with a multicycle conference. We first derived an achievable rate for the general channel, and then we specialized it to the single-cycle conference for which we obtained an explicit achievable rate. This rate is superior to the explicit expression that can be obtained with joint decoding.

As for the future work, we focus on the Gaussian relay channel with coded modulation. The objective is to gain analytic insight into the EAF rates by finding analytic approximations to the integrals used in the evaluation. This is necessary in order to compare the HD-EAF and DHD methods and find analytically the regions in which each of the two is superior. Analytic approximations are also necessary to characterize the relationship between the SNR on the source–relay link and the minimal cardinality of $\hat{Y}_1$ that achieves the maximum TS-DHD rate.

Appendix A
Proof of the Equivalence of Joint Decoding and the Classic EAF for Section II-C

Let us now compare the rates obtained with joint decoding (Proposition 2) with the rates obtained with the sequential decoding of [2, Th. 6]: to that end, we consider the joint-decoding result of Proposition 2 with the extended probability chain of (7)

$$p(x,x_1,y,y_1,\hat{y}_1,\hat{\hat{y}}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)p(\hat{\hat{y}}_1|\hat{y}_1)$$

where $\hat{\hat{Y}}_1$ represents the information relayed to the destination. Expanding the expressions of Proposition 2 using the assignment (10), similarly to Proposition 1, we obtain the expressions

$$R \leq I(X;Y|X_1) + \min\{qI(X;\hat{Y}_1|X_1,Y), \\ I(X_1;Y) - qI(\hat{Y}_1;Y_1|X,X_1,Y)\} \tag{A.1}$$

subject to

$$I(X_1;Y) \geq qI(\hat{Y}_1;Y_1|X,X_1,Y) \\ = q(I(\hat{Y}_1;Y_1|X_1,Y) - I(X;\hat{Y}_1|X_1,Y)). \tag{A.2}$$

We now make the following observations.

1) Setting $q = 1$, we obtain Proposition 2. Additionally, if $I(X_1;Y) > I(\hat{Y}_1;Y_1|X_1,Y)$, then both Proposition 2 and [2, Th. 6] give identical expressions.
2) When $q = 1$ and

$$I(\hat{Y}_1;Y_1|X_1,Y) - I(X;\hat{Y}_1|X_1,Y) < I(X_1;Y) \\ < I(\hat{Y}_1;Y_1|X_1,Y) \tag{A.3}$$

then for the same mapping $p(\hat{y}_1|x_1,y_1)$, we see that Proposition 2 provides rate but [2, Th. 6] does not. The rate expression under these conditions is

$$R \leq I(X;Y|X_1) + I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y). \tag{A.4}$$

3) Now, fix the probability chain $p(x)p(x_1)$ $p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)$ and examine the expressions (A.1) and (A.2) when (A.3) holds. When $q < 1$, then (A.3) guarantees that condition (A.2) is still satisfied. If q is close enough to 1 such that we also have $I(X_1;Y) \leq qI(\hat{Y}_1;Y_1|X_1,Y)$, the rate from (A.1), i.e.,

$$R \leq I(X;Y|X_1) + I(X_1;Y) - qI(\hat{Y}_1;Y_1|X,X_1,Y)$$

is now greater than (A.4). In this case, we can keep decreasing q until

$$I(X_1;Y) - qI(\hat{Y}_1;Y_1|X,X_1,Y) = qI(X;\hat{Y}_1|X_1,Y) \tag{A.5}$$

at which point the rate becomes

$$R \leq I(X;Y|X_1) + qI(X;\hat{Y}_1|X_1,Y). \tag{A.6}$$

This rate can be obtained from [2, Th. 6] by applying the extended probability chain of (7), as long as $I(X_1;Y) \geq qI(\hat{Y}_1,Y_1|X_1,Y)$.

We, therefore, conclude that all the rates that joint decoding allows can also be obtained or exceeded by the original EAF with an appropriate TS.[5]

[5]This argument is due to Shlomo Shamai and Gerhard Kramer.

Note that equality in (A.5) implies

$$q_{\text{opt}} = \min\left\{1, \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X,X_1,Y) + I(X;\hat{Y}_1|X_1,Y)}\right\} \\ = \min\left\{1, \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X_1,Y)}\right\}$$

hence q_{opt} is the maximum q that makes the mapping $p(\hat{y}_1|x_1,y_1)$ feasible for [2, Th. 6]. Plugging q_{opt} into (A.6), we obtain the rate expression of Proposition 1.

Finally, consider again the region where joint decoding is useful (A.3)

$$I(\hat{Y}_1;Y_1|X,X_1,Y) \leq I(X_1;Y) \leq I(\hat{Y}_1;Y_1|X_1,Y) \\ \Rightarrow 0 \leq I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y) \\ \leq I(\hat{Y}_1;Y_1|X_1,Y) - I(\hat{Y}_1;Y_1|X,X_1,Y) \\ \Rightarrow 0 \leq I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y) \leq I(X;\hat{Y}_1|X_1,Y) \\ \Rightarrow 0 \leq \frac{I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)}{I(X;\hat{Y}_1|X_1,Y)} \leq 1.$$

If $I(X;\hat{Y}_1|X_1,Y) > 0$, then using TS on $\hat{Y}_1$ with

$$q = \frac{I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)}{I(X;\hat{Y}_1|X_1,Y)} \tag{A.7}$$

in (11) and (12) yields

$$I(X;Y|X_1) + qI(X;\hat{Y}_1|X_1,Y) \\ = I(X;Y|X_1) + I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)$$

as long as $I(X_1;Y) \geq qI(\hat{Y}_1;Y_1|X_1,Y)$, or equivalently

$$q \leq \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X_1,Y)}. \tag{A.8}$$

Plugging assignment (A.7) into (A.8), we obtain

$$\frac{I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)}{I(X;\hat{Y}_1|X_1,Y)} \leq \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X_1,Y)} \\ \Rightarrow \Big(I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)\Big) I(\hat{Y}_1;Y_1|X_1,Y) \\ \leq I(X_1;Y)I(X;\hat{Y}_1|X_1,Y) \\ \Rightarrow I(X_1;Y)I(\hat{Y}_1;Y_1|X_1,Y) - I(X_1;Y)I(X;\hat{Y}_1|X_1,Y) \\ \leq I(\hat{Y}_1;Y_1|X,X_1,Y)I(\hat{Y}_1;Y_1|X_1,Y) \\ \Rightarrow I(X_1;Y)I(\hat{Y}_1;Y_1|X,X_1,Y) \\ \leq I(\hat{Y}_1;Y_1|X,X_1,Y)I(\hat{Y}_1;Y_1|X_1,Y) \\ \Rightarrow I(X_1;Y) \leq I(\hat{Y}_1;Y_1|X_1,Y)$$

as long as $I(\hat{Y}_1;Y_1|X,X_1,Y) > 0$, which is the region where joint decoding is supposed to be useful. Hence, the joint-decoding rate of Proposition 2 can be obtained by TS on the [2, Th. 6] expression. Therefore, joint decoding does not improve on the rate of [2, Th. 6]. In fact, the rate of Proposition 1 is always at least as large as that of Proposition 2.

APPENDIX B
EXPRESSIONS FOR SECTION III

A. Hard-Decision Estimate-and-Forward

We evaluate $I(X;\hat{Y}_1,Y)$, with $p(\hat{Y}_1|Y_1)$ given by (19a) and (19b) using

$$I(X;\hat{Y}_1,Y)=I(X;\hat{Y}_1)+I(X;Y|\hat{Y}_1).$$

1) Evaluating $I(X;\hat{Y}_1)$: note that both X and $\hat{Y}_1$ are discrete RVs, therefore $I(X;\hat{Y}_1)$ can be evaluated using the discrete entropies. The conditional distribution of $\hat{Y}_1$ given X is given by

$$p(\hat{Y}_1|X=\sqrt{P})=\begin{cases}P_1\cdot P_{\text{no erase}}, & 1\\ 1-P_{\text{no erase}}, & E\\ (1-P_1)P_{\text{no erase}}, & -1\end{cases}\quad\text{(B.1)}$$

where

$$P_1=\Pr(Y_1>0|X=\sqrt{P}).$$

$p(\hat{Y}_1|X=-\sqrt{P})$ can be obtained from $p(\hat{Y}_1|X=\sqrt{P})$ by switching 1 and -1 in (B.1).

2) Evaluating $I(X;Y|\hat{Y}_1)$: write first

$$I(X;Y|\hat{Y}_1)=h(Y|\hat{Y}_1)-h(Y|\hat{Y}_1,X)$$

and we note that

$$\begin{aligned}h(Y|\hat{Y}_1,X)&=h(X+N|\hat{Y}_1,X)\\&=h(N|\hat{Y}_1,X)\\&=h(N)=\frac{1}{2}\log_2(2\pi e\sigma^2).\end{aligned}$$

Using the chain rule, we write

$$\begin{aligned}h(Y|\hat{Y}_1)=&\,p(\hat{Y}_1=1)h(Y|\hat{Y}_1=1)\\&+p(\hat{Y}_1=E)h(Y|\hat{Y}_1=E)\\&+p(\hat{Y}_1=-1)h(Y|\hat{Y}_1=-1)\end{aligned}$$

$p(\hat{Y}_1)$ can be obtained by combining (17) and (B.1), which results in

$$p(\hat{Y}_1)=\begin{cases}\frac{1}{2}P_{\text{no erase}}, & 1\\ 1-P_{\text{no erase}}, & E\\ \frac{1}{2}P_{\text{no erase}}, & -1\end{cases}\quad\text{(B.2)}$$

and we note that $h(Y|\hat{Y}_1=E)=h(Y)$, because erasure is equivalent to no prior information. Finally, we note that by definition

$$\begin{aligned}h(Y)&=-\int_{y=-\infty}^{\infty}f(y)\log_2(f(y))dy,\\ f(Y)&=\Pr(X=\sqrt{P})f(Y|X=\sqrt{P})\\&\quad+\Pr(X=-\sqrt{P})f(Y|X=-\sqrt{P})\\&=\frac{1}{2}(G_y(\sqrt{P},\sigma^2)+G_y(-\sqrt{P},\sigma^2))\end{aligned}\quad\text{(B.3)}$$

where

$$G_x(a,b)=\frac{1}{\sqrt{2\pi b}}e^{-\frac{(x-a)^2}{2b}}.$$

Next, we have

$$\begin{aligned}&h(Y|\hat{Y}_1=1)\\&\quad=-\int_{y=-\infty}^{\infty}f(y|\hat{y}_1=1)\log_2(f(y|\hat{y}_1=1))dy\quad\text{(B.4)}\\&f(Y|\hat{Y}_1=1)\\&\quad=\frac{f(Y,\hat{Y}_1=1)}{\Pr(\hat{Y}_1=1)}\\&\quad=\frac{f(Y,Y_1>0)P_{\text{no erase}}}{\Pr(Y_1>0)P_{\text{no erase}}}\\&\quad=\frac{f(Y,Y_1>0)}{\Pr(Y_1>0)}\quad\text{(B.5)}\\&f(Y,Y_1>0)\\&\quad=\Pr(X=\sqrt{P})f(Y,Y_1>0|X=\sqrt{P})\\&\qquad+\Pr(X=-\sqrt{P})f(Y,Y_1>0|X=-\sqrt{P})\\&\quad=\frac{1}{2}(f(Y,Y_1>0|X=\sqrt{P})\\&\qquad+f(Y,Y_1>0|X=-\sqrt{P})).\quad\text{(B.6)}\end{aligned}$$

Using

$$\begin{aligned}f_{Y,Y_1}(y,y_1|x)&=\mathcal{N}\left(\begin{pmatrix}x\\ g\cdot x\end{pmatrix},\begin{pmatrix}\sigma^2&0\\0&\sigma_1^2\end{pmatrix}\right)\\&=G_y(x,\sigma^2)G_{y_1}\left(g\cdot x,\sigma_1^2\right)\end{aligned}$$

we obtain

$$\begin{aligned}f(Y,Y_1>0|X)&=\int_{y_1=0}^{\infty}f(y,y_1|x)dy_1\\&=G_y(x,\sigma^2)\int_{y_1=0}^{\infty}G_{y_1}\left(g\cdot x,\sigma_1^2\right)dy_1.\end{aligned}$$

Next, we need to evaluate $I(\hat{Y}_1;Y_1|Y)=h(Y_1|Y)-h(Y_1|Y,\hat{Y}_1)$.

1) $h(Y_1|Y)=h(Y,Y_1)-h(Y)$. Here

$$\begin{aligned}&h(Y,Y_1)\\&\quad=-\int_{y=-\infty}^{\infty}\int_{y_1=-\infty}^{\infty}f(y,y_1)\log_2(f(y,y_1))dy\,dy_1\\&f(Y,Y_1)\\&\quad=\frac{1}{2}(f(Y,Y_1|X=\sqrt{P})+f(Y,Y_1|X=-\sqrt{P}))\\&f(Y,Y_1|X)\\&\quad=G_y(x,\sigma^2)G_{y_1}\left(g\cdot x,\sigma_1^2\right).\end{aligned}$$

2) By the definition of conditional entropy, we have

$$\begin{aligned}h(Y_1|Y,\hat{Y}_1)=&\,p(\hat{Y}_1=1)h(Y_1|Y,\hat{Y}_1=1)\\&+p(\hat{Y}_1=E)h(Y_1|Y,\hat{Y}_1=E)\\&+p(\hat{Y}_1=-1)h(Y_1|Y,\hat{Y}_1=-1)\end{aligned}$$

where $h(Y_1|Y,\hat{Y}_1=E)=h(Y_1|Y)$, and for $\hat{Y}_1=1$, for example, we have

$$\begin{aligned}h(Y_1|Y,\hat{Y}_1=1)\\ &=-\int_{y=-\infty}^{\infty}\int_{y_1=-\infty}^{\infty}(f(y,y_1|\hat{y}_1=1)\\ &\qquad\times\log_2(f(y_1|y,\hat{y}_1=1)))dydy_1.\end{aligned}$$

Finally, we need to derive the distributions $f(y,y_1|\hat{y}_1=1)$ and $f(y_1|y,\hat{y}_1=1)$. Begin with

$$\begin{aligned}f_{Y,Y_1|\hat{Y}_1}(y,y_1|\hat{y}_1=1)&=\frac{f_{Y,Y_1,\hat{Y}_1}(y,y_1,\hat{y}_1=1)}{\Pr(\hat{y}_1=1)}\\ &=\frac{f_{Y,Y_1,\hat{Y}_1}(y,y_1,y_1>0)P_{\text{no erase}}}{\Pr(y_1>0)P_{\text{no erase}}}\\ &=f(y,y_1|y_1>0)\\ &=\begin{cases}\frac{f_{Y,Y_1}(y,y_1)}{\Pr(Y_1>0)}, & y_1>0\\ 0, & y_1\le 0\end{cases}\end{aligned}$$

and due to the symmetry, $\Pr(Y_1>0)=\Pr(Y_1\le 0)=\frac{1}{2}$. We also have

$$\begin{aligned}f(Y_1|Y,\hat{Y}_1=1)&=\frac{f(Y_1,Y|\hat{Y}_1=1)}{f(Y|\hat{Y}_1=1)}\\ &=\frac{f(Y_1,Y|Y_1>0)}{f(Y|Y_1>0)}\\ &=\frac{\frac{f(Y_1,Y)}{\Pr(Y_1>0)}}{\frac{f(Y,Y_1>0)}{\Pr(Y_1>0)}}\\ &=\frac{f(Y_1,Y)}{f(Y,Y_1>0)},\qquad Y_1>0\\ f(Y_1|Y,\hat{Y}_1=1)&=0,\qquad Y_1\le 0.\end{aligned}$$

B. Evaluation of the Rate With DHD

We evaluate the achievable rate using $I(X;Y,\hat{Y}_1)=I(X;\hat{Y}_1)+I(X;Y|\hat{Y}_1)$. The distribution of $\hat{Y}_1$ is given by

$$\begin{aligned}&\Pr(\hat{Y}_1=1)\\ &\ =\Pr(Y_1>T)\\ &\ =\frac{1}{2}(\Pr(Y_1>T|X=\sqrt{P})+\Pr(Y_1>T|X=-\sqrt{P}))\\ &\ =\frac{1}{2}\left(\int_{y_1>T}G_{y_1}(g\sqrt{P},\sigma_1^2)dy_1+\int_{y_1>T}G_{y_1}(-g\sqrt{P},\sigma_1^2)dy_1\right)\\ &\Pr(\hat{Y}_1=E)\\ &\ =\Pr(|Y_1|\le T)\\ &\ =\frac{1}{2}(\Pr(|Y_1|\le T|X=\sqrt{P})+\Pr(|Y_1|\le T|X=-\sqrt{P}))\\ &\ =\frac{1}{2}\left(\int_{y_1=-T}^{T}G_{y_1}(g\sqrt{P},\sigma_1^2)dy_1\right.\\ &\qquad\left.+\int_{y_1=-T}^{T}G_{y_1}(-g\sqrt{P},\sigma_1^2)dy_1\right)\end{aligned}$$

and by symmetry, $\Pr(\hat{Y}_1=1)=\Pr(\hat{Y}_1=-1)$ and $H(\hat{Y}_1|X=\sqrt{P})=H(\hat{Y}_1|X=-\sqrt{P})$. Therefore, we need the conditional distribution $p(\hat{Y}_1|X=\sqrt{P})$

$$\begin{aligned}\Pr(\hat{Y}_1=1|X=\sqrt{P})&=\Pr(Y_1>T|X=\sqrt{P})\\ &=\int_{y_1>T}G_{y_1}(g\sqrt{P},\sigma_1^2)dy_1\\ \Pr(\hat{Y}_1=-1|X=\sqrt{P})&=\Pr(Y_1<-T|X=\sqrt{P})\\ &=\int_{y_1<-T}G_{y_1}(g\sqrt{P},\sigma_1^2)dy_1\\ \Pr(\hat{Y}_1=E|X=\sqrt{P})&=1-\Pr(\hat{Y}_1=1|X=\sqrt{P})\\ &\quad-\Pr(\hat{Y}_1=-1|X=\sqrt{P}).\end{aligned}$$

This allows us to evaluate $I(X;\hat{Y}_1)=H(\hat{Y}_1)-H(\hat{Y}_1|X)$. For evaluating $I(X;Y|\hat{Y}_1)$, note that

$$\begin{aligned}h(Y|\hat{Y}_1,X)&=h(X+N|\hat{Y}_1,X)\\ &=h(N|\hat{Y}_1,X)\\ &=h(N)=\frac{1}{2}\log_2(2\pi e\sigma^2)\end{aligned}$$

and we need only to evaluate $h(Y|\hat{Y}_1)$. By definition

$$\begin{aligned}h(Y|\hat{Y}_1)=&\Pr(\hat{Y}_1=1)h(Y|\hat{Y}_1=1)\\ &+\Pr(\hat{Y}_1=E)h(Y|\hat{Y}_1=E)\\ &+\Pr(\hat{Y}_1=-1)h(Y|\hat{Y}_1=-1)\end{aligned}$$

and note that $h(Y|\hat{Y}_1=E)=h(Y)$. Finally, $h(Y|\hat{Y}_1=1)$ can be obtained using (B.7a)–(B.7c) shown at the bottom of the following page. Evaluating $I(\hat{Y}_1;Y_1|Y)$, we have

$$\begin{aligned}I(\hat{Y}_1;Y_1|Y)&=H(\hat{Y}_1|Y)-H(\hat{Y}_1|Y,Y_1)\\ &\overset{(a)}{=}H(\hat{Y}_1|Y)\\ &=H(\hat{Y}_1)+h(Y|\hat{Y}_1)-h(Y)\end{aligned}$$

where (a) is due to the deterministic mapping from Y_1 to $\hat{Y}_1$, and $h(Y)$ can be evaluated using (B.3).

1) DHD When $T\to 0$: As $T\to 0$, we have that $\Pr(\hat{Y}_1=E)\to 0$ and $\hat{Y}_1$ converges in distribution to a Bernoulli RV with probability $1/2$. Therefore

$$\begin{aligned}&f(Y,\hat{Y}_1=1)\\ &\quad=\frac{1}{2}\Big(G_y(\sqrt{P},\sigma^2)\Pr(Y_1>T|X=\sqrt{P})\\ &\qquad\quad+G_y(-\sqrt{P},\sigma^2)\Pr(Y_1>T|X=-\sqrt{P})\Big)\\ &\quad\overset{T\to 0}{\approx}\frac{1}{2}\Big(G_y(\sqrt{P},\sigma^2)\Pr(Y_1>0|X=\sqrt{P})\\ &\qquad\quad+G_y(-\sqrt{P},\sigma^2)\Pr(Y_1>0|X=-\sqrt{P})\Big)\\ &\quad=\frac{1}{2}(G_y(\sqrt{P},\sigma^2)P_++G_y(-\sqrt{P},\sigma^2)(1-P_+))\end{aligned}$$

where $P_+=\Pr(Y_1>0|X=\sqrt{P})$. Now, letting $g\to 0$, we have that $P_+\to\frac{1}{2}$, and therefore

$$\begin{aligned}&f(Y|\hat{Y}_1=1)\xrightarrow{g\to 0,T\to 0}f(Y)\\ \Rightarrow\ &h(Y|\hat{Y}_1=1)\xrightarrow{g\to 0,T\to 0}h(Y).\end{aligned}$$

We conclude that as $g \to 0, T \to 0$, then $h(Y|\hat{Y}_1) \to h(Y)$, and therefore, the $I(Y_1;\hat{Y}_1|Y)$ becomes

$$I(Y_1;\hat{Y}_1|Y) = H(\hat{Y}_1) + h(Y|\hat{Y}_1) - h(Y) \xrightarrow{g\to 0, T\to 0} 1.$$

Using the continuity of $I(Y_1;\hat{Y}_1|Y)$, we conclude that for small values of g, as T decreases, then $I(Y_1;\hat{Y}_1|Y)$ is bounded from below. This implies that for small g and small C, feasibility is obtained only for large T, which in turn implies low rate.

C. Evaluating the Information Rate With TS-DHD

1) Evaluating $I(X;Y,\hat{Y}_1)$: We first write

$$I(X;Y,\hat{Y}_1) = I(X;\hat{Y}_1) + I(X;Y|\hat{Y}_1).$$

Evaluating $I(X;\hat{Y}_1) = H(\hat{Y}_1) - H(\hat{Y}_1|X)$ requires the marginal of $\hat{Y}_1$. Using the mapping defined in (21), we find the marginal distribution of $\hat{Y}_1$

$$\Pr(\hat{Y}_1) = \begin{cases} 1, & (1-P_{\text{erase}})\Pr(Y_1 > T) \\ E, & \Pr(|Y_1| \le T) + P_{\text{erase}}\Pr(|Y_1| > T) \\ -1, & (1-P_{\text{erase}})\Pr(Y_1 < -T) \end{cases}$$

where

$$\begin{aligned}\Pr(Y_1 > T) &= \Pr(Y_1 < -T) \\ &= \frac{1}{2}\int_{y_1=T}^{\infty}[G_{y_1}(\sqrt{P},\sigma_1^2) + G_{y_1}(-\sqrt{P},\sigma_1^2)]dy_1 \\ \Pr(|Y_1| < T) &= \frac{1}{2}\int_{y_1=-T}^{T}[G_{y_1}(\sqrt{P},\sigma_1^2) + G_{y_1}(-\sqrt{P},\sigma_1^2)]dy_1.\end{aligned}$$

Also, due to the symmetry, we have that $H(\hat{Y}_1|X=\sqrt{P}) = H(\hat{Y}_1|X=-\sqrt{P})$, and therefore, we need only to find the conditional $\Pr(\hat{Y}_1|X=\sqrt{P})$, which is shown in (B.8) at the bottom of the page, and we note that $f_{Y_1|X}(y_1|x=\sqrt{P}) = G_{y_1}(\sqrt{P},\sigma_1^2)$.

Next, we need to evaluate $I(X;Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(Y|\hat{Y}_1,X)$. We first note that

$$\begin{aligned}h(Y|\hat{Y}_1,X) &= h(X+N|X,\hat{Y}_1) \\ &= h(N|X,\hat{Y}_1) \\ &= h(N) = \frac{1}{2}\log_2\left(2\pi e\sigma_1^2\right).\end{aligned}$$

Last, we have

$$\begin{aligned}h(Y|\hat{Y}_1) = &\Pr(\hat{Y}_1=1)h(Y|\hat{Y}_1=1) \\ &+ \Pr(\hat{Y}_1=E)h(Y|\hat{Y}_1=E) \\ &+ \Pr(\hat{Y}_1=-1)h(Y|\hat{Y}_1=-1).\end{aligned}$$

We note that $h(Y|\hat{Y}_1=E) = h(Y)$ and that $h(Y|\hat{Y}_1=1)$ and $h(Y|\hat{Y}_1=-1)$ are calculated exactly as in part B of Appendix B for the DHD case.

2) Evaluating $I(\hat{Y}_1;Y_1|Y)$: Begin by writing

$$\begin{aligned}I(\hat{Y}_1;Y_1|Y) &= h(\hat{Y}_1|Y) - h(\hat{Y}_1|Y_1,Y) \\ &= h(Y|\hat{Y}_1) + H(\hat{Y}_1) - h(Y) - h(\hat{Y}_1|Y_1)\end{aligned}$$

where we used the fact that given Y_1, $\hat{Y}_1$ is independent of Y. All the terms in the above expressions have been calculated in the previous section, except $h(\hat{Y}_1|Y_1)$

$$\begin{aligned}h(\hat{Y}_1|Y_1) = &\Pr(\hat{Y}_1 > T)h(\hat{Y}_1|Y_1 > T) \\ &+ \Pr(|Y_1| \le T)h(\hat{Y}_1|\,|Y_1| \le T) \\ &+ \Pr(Y_1 < -T)h(\hat{Y}_1|Y_1 < -T) \\ = &\Pr(\hat{Y}_1 > T)H(P_{\text{erase}}, 1-P_{\text{erase}}) \\ &+ \Pr(\hat{Y}_1 < -T)H(P_{\text{erase}}, 1-P_{\text{erase}}) \\ = &(1 - P(|Y_1| \le T)H(P_{\text{erase}}, 1-P_{\text{erase}}).\end{aligned}$$

D. Gaussian-Quantization Estimate-and-Forward

Here the relay uses the assignment of (15)

$$\hat{Y}_1 = Y_1 + N_Q, \quad N_Q \sim \mathcal{N}\left(0,\sigma_Q^2\right).$$

$$h(Y|\hat{Y}_1=1) = -\int_{y=-\infty}^{\infty} f(y|\hat{y}_1=1)\log_2(f(y|\hat{y}_1=1))dy \tag{B.7a}$$

$$\begin{aligned}f_{Y|\hat{Y}_1}(y|\hat{y}_1=1) &= f(y|y_1>T) \\ &= \frac{f(y,y_1>T)}{\Pr(Y_1>T)}\end{aligned} \tag{B.7b}$$

$$\begin{aligned}f_{Y,Y_1}(y,y_1>T) &= \frac{1}{2}\Big(f(y,y_1>T|X=\sqrt{P}) + f(y,y_1>T|X=-\sqrt{P})\Big) \\ &= \frac{1}{2}\Big(G_y(\sqrt{P},\sigma^2)\Pr(Y_1>T|X=\sqrt{P}) + G_y\left(-\sqrt{P},\sigma^2\right)\Pr(Y_1>T|X=-\sqrt{P})\Big).\end{aligned} \tag{B.7c}$$

$$\Pr(\hat{Y}_1|X=\sqrt{P}) = \begin{cases} 1, & (1-P_{\text{erase}})\Pr(Y_1>T|X=\sqrt{P}) \\ E, & \Pr(|Y_1| \le T|X=\sqrt{P}) + P_{\text{erase}}\Pr(|Y_1|>T|X=\sqrt{P}) \\ -1, & (1-P_{\text{erase}})\Pr(Y_1<-T|X=\sqrt{P}) \end{cases} \tag{B.8}$$

We first evaluate

$$I(X;Y,\hat{Y}_1) = h(Y,\hat{Y}_1) - h(Y,\hat{Y}_1|X).$$

1) For $h(Y,\hat{Y}_1)$, see (B.9) shown at the bottom of the page.
2) We also have

$$\begin{aligned} h(Y,\hat{Y}_1|X) &= h(X+N, gX+N_1+N_Q|X) \\ &= h(N, N_1+N_Q|X) \\ &= h(N) + h(N_1+N_Q) \\ &= \frac{1}{2}\log_2\left((2\pi e)^2\sigma^2\left(\sigma_1^2+\sigma_Q^2\right)\right). \end{aligned}$$

Last, we need to evaluate

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

where

$$\begin{aligned} h(\hat{Y}_1|Y_1,Y) &= h(Y_1+N_Q|Y_1,Y) \\ &= h(N_Q|Y_1,Y) \\ &= h(N_Q) = \frac{1}{2}\log_2(2\pi e\sigma_Q^2). \end{aligned}$$

E. Approximation of HD-EAF for $\sigma^2 \to \infty$

Now, let us evaluate the rate for HD-EAF as the SNR goes to zero. From (14a)

$$R \le I(X;Y,\hat{Y}_1) = I(X;\hat{Y}_1) + I(X;Y|\hat{Y}_1)$$

and

$$\begin{aligned} I(X;Y|\hat{Y}_1) &= h(Y|\hat{Y}_1) - h(Y|X,\hat{Y}_1) \\ &= \Pr(\hat{Y}_1=1)h(Y|\hat{Y}_1=1) \\ &\quad + \Pr(\hat{Y}_1=E)h(Y|\hat{Y}_1=E) \\ &\quad + \Pr(\hat{Y}_1=-1)h(Y|\hat{Y}_1=-1) - h(N). \end{aligned}$$

Using (B.4)–(B.6), we have

$$\begin{aligned} &h(Y|\hat{Y}_1=1) \\ &\quad = -\int_{y=-\infty}^{\infty}\left(f_{Y|\hat{Y}_1}(y|\hat{y}_1=1)\log_2\left(f_{Y|\hat{Y}_1}(y|\hat{y}_1=1)\right)\right)dy \\ &f_{Y|\hat{Y}_1}(y|\hat{y}_1=1) \\ &\quad = \frac{f_{Y,Y_1}(y,y_1>0)P_{\text{no erase}}}{\Pr(Y_1>0)P_{\text{no erase}}} \\ &\quad = \frac{f_{Y,Y_1}(y,y_1>0)}{\Pr(Y_1>0)} \\ &f_{Y,Y_1}(y,y_1>0) \\ &\quad = \frac{1}{2}\Big(f_{Y,Y_1|X}(y,y_1>0|x=\sqrt{P}) \\ &\qquad + f_{Y,Y_1|X}(y,y_1>0|x=-\sqrt{P})\Big) \\ &\quad = \frac{1}{2}\Big(G_y(\sqrt{P},\sigma^2)\Pr(Y_1>0|X=\sqrt{P}) \\ &\qquad + G_y(-\sqrt{P},\sigma^2)(1-\Pr(Y_1>0|X=\sqrt{P}))\Big) \\ &\quad = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\Big(\frac{1}{2}e^{\frac{y\sqrt{P}}{\sigma^2}}\Pr(Y_1>0|X=\sqrt{P}) \\ &\qquad + \frac{1}{2}e^{-\frac{y\sqrt{P}}{\sigma^2}}(1-\Pr(Y_1>0|X=\sqrt{P}))\Big)e^{-\frac{P}{2\sigma^2}} \\ &\quad = \frac{e^{-\frac{P}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\frac{1}{2}\left(\left(\frac{1}{2}-\delta\right)e^{\frac{y\sqrt{P}}{\sigma^2}} + \left(\frac{1}{2}+\delta\right)e^{-\frac{y\sqrt{P}}{\sigma^2}}\right) \\ &\quad = \frac{e^{-\frac{P}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\left(\frac{1}{2}\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right) - \delta\sinh\left(\frac{y\sqrt{P}}{\sigma^2}\right)\right) \\ &\quad \overset{(a)}{\approx} \frac{1}{2}G_y(0,\sigma^2) \end{aligned}$$

when $\sigma^2 \to \infty$ and $\delta \in \left[-\frac{1}{2},\frac{1}{2}\right]$ is selected such that $\Pr(Y_1 > 0|X=\sqrt{P}) = \frac{1}{2}-\delta$. The approximation in (a) is because for small $|y|$, $\sinh\left(\frac{y\sqrt{P}}{\sigma^2}\right) \approx 0$ and $\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right) \approx 1$, and for large $|y|$, both $e^{-\frac{y^2}{2\sigma^2}}\sinh\left(\frac{y\sqrt{P}}{\sigma^2}\right)$ and $e^{-\frac{y^2}{2\sigma^2}}\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right)$ behave as $e^{-\frac{y^2}{2\sigma^2}}$.

Hence

$$\begin{aligned} &h(Y|\hat{Y}_1=1) \\ &\quad \approx -\int_{y=-\infty}^{\infty}\frac{G_y(0,\sigma^2)}{2\Pr(Y_1>0)}\log_2\left(\frac{G_y(0,\sigma^2)}{2\Pr(Y_1>0)}\right)dy \\ &\quad = -\frac{1}{2\Pr(Y_1>0)} \\ &\qquad \int_{y=-\infty}^{\infty}G_y(0,\sigma^2)\Big[\log_2\left(G_y(0,\sigma^2)\right) \\ &\qquad\qquad - \log_2\left(2\Pr(Y_1>0)\right)\Big]dy \\ &\quad = \frac{1}{2\Pr(Y_1>0)}\left[h(N) + \log_2\left(2\Pr(Y_1>0)\right)\right] \end{aligned}$$

and using $\Pr(Y_1>0) = \Pr(Y_1\le 0) = \frac{1}{2}$ and $h(Y|\hat{Y}_1=1) = h(Y|\hat{Y}_1=-1)$, we obtain

$$\begin{aligned} h(Y|\hat{Y}_1) &\approx \frac{1}{2}P_{\text{no erase}}h(N) + (1-P_{\text{no erase}})h(N) \\ &\quad + \frac{1}{2}P_{\text{no erase}}h(N) \\ &= h(N). \end{aligned}$$

Therefore, at low SNR, Y and $\hat{Y}_1$ become independent. Then, $I(X;Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(N) \approx 0$ and the information rate

$$\begin{aligned} h(Y,\hat{Y}_1) &= -\int_{\hat{y}_1=-\infty}^{\infty}\int_{y=-\infty}^{\infty}\left(f_{Y,\hat{Y}_1}(y,\hat{y}_1)\log_2(f_{Y,\hat{Y}_1}(y,\hat{y}_1))\right)dyd\hat{y}_1 \\ f_{Y,\hat{Y}_1}(y,\hat{y}_1) &= \frac{1}{2}\left(G_y\left(\sqrt{P},\sigma^2\right)G_{\hat{y}_1}\left(g\sqrt{P},\sigma_1^2+\sigma_Q^2\right) + G_y\left(-\sqrt{P},\sigma^2\right)G_{\hat{y}_1}\left(-g\sqrt{P},\sigma_1^2+\sigma_Q^2\right)\right). \end{aligned} \qquad \text{(B.9)}$$

becomes [see (B.1) and (B.2)]

$$
\begin{aligned}
R &\le I(X;\hat{Y}_1)\\
&= H(\hat{Y}_1) - H(\hat{Y}_1|X)\\
&= H\left(\frac{1}{2}P_{\text{no erase}}, 1-P_{\text{no erase}}, \frac{1}{2}P_{\text{no erase}}\right)\\
&\qquad - H\left(P_1 P_{\text{no erase}}, 1-P_{\text{no erase}}, (1-P_1)P_{\text{no erase}}\right)\\
&= P_{\text{noerase}}(1 + P_1\log_2(P_1) + (1-P_1)\log_2(1-P_1))\\
&= P_{\text{no erase}}(1 - H(P_1, 1-P_1))
\end{aligned}
$$

where $H(\mathbf{p})$ is the discrete entropy for the specified pdf $\mathbf{p}$, and $P_1 = \Pr(Y_1 > 0|X = \sqrt{P})$. Now, consider the feasibility condition $C \ge I(Y_1;\hat{Y}_1|Y)$

$$
\begin{aligned}
I(Y_1;\hat{Y}_1|Y) &= H(\hat{Y}_1|Y) - H(\hat{Y}_1|Y_1,Y)\\
&\overset{(a)}{\approx} H(\hat{Y}_1) - H(\hat{Y}_1|Y_1)\\
&= H\left(\frac{1}{2}P_{\text{no erase}}, 1-P_{\text{no erase}}, \frac{1}{2}P_{\text{no erase}}\right)\\
&\quad - H(P_{\text{no erase}}, 1-P_{\text{no erase}})\\
&= -2\cdot\frac{1}{2}P_{\text{no erase}}\log_2\left(\frac{1}{2}P_{\text{no erase}}\right)\\
&\quad - (1-P_{\text{no erase}})\log_2(1-P_{\text{no erase}})\\
&\quad + P_{\text{no erase}}\log_2(P_{\text{no erase}})\\
&\quad + (1-P_{\text{no erase}})\log_2(1-P_{\text{no erase}})\\
&= P_{\text{no erase}},
\end{aligned}
$$

where in (a) we used the fact that $\hat{Y}_1$ and Y are independent as $\sigma^2 \to \infty$, and that given Y_1, $\hat{Y}_1$ is independent of Y. Therefore, for low SNR, we set $P_{\text{no erase}} = \min\{C,1\}$ and the rate becomes

$$
R_{HD-EAF} \le \min\{C,1\}(1 - H(P_1, 1-P_1)). \qquad \text{(B.10)}
$$

F. Derivation for (23)

$$
\begin{aligned}
R &\le I(X;Y,\hat{Y}_1)\\
&= h(Y,\hat{Y}_1) - h(Y,\hat{Y}_1|X)\\
&= h(Y) + h(\hat{Y}_1) - h(N, N_1+N_Q|X)\\
&= h(Y) - h(N|X) + h(\hat{Y}_1) - h(N_1+N_Q|X)\\
&= I(X;Y) + I(X;\hat{Y}_1)\\
&\approx I(X;\hat{Y}_1)\\
&= h(\hat{Y}_1) - h(N_1+N_Q).
\end{aligned}
$$

APPENDIX C
PROOF OF COROLLARY 2

In the following, we highlight only the modifications from the general broadcast result due to the application of DAF at the last conference step from Rx_2 to Rx_1, and the fact that we transmit a single message.

1) Codebook Generation and Encoding at the Transmitter: The transmitter generates 2^{nR} codewords $\mathbf{x}$ in an i.i.d. manner according to $p(\mathbf{x}(w)) = \prod_{i=1}^n p_X(x_i)$, $w \in \mathcal{W} = \{1,2,\ldots,2^{nR}\}$. For transmission of the message w_i at time i, the transmitter outputs $\mathbf{x}(w_i)$.

2) Codebook Generation at the Rx_1*:* The K conference steps from Rx_1 to Rx_2 are carried out exactly as in Section IV-B4. The first $K-1$ steps from Rx_2 to Rx_1 are carried out as in Section IV-B5. The Kth conference step from Rx_2 to Rx_1 is different from that of Theorem 3, as after the Kth step from Rx_1 to Rx_2, Rx_2 may decode the message because Rx_2 received all the K conference messages from Rx_1. Then, Rx_2 uses DAF for its Kth conference transmission to Rx_1. Therefore, Rx_2 simply partitions $\mathcal{W}$ into $2^{n\alpha C_{21}}$ subsets in a uniform and independent manner.

3) Encoding and Decoding at the Kth Conference Step From Rx_2 *to* Rx_1*:*

- Before the Kth conference step, Rx_2 decodes its message using his channel output and all the K conference messages received from Rx_1. This can be done with an arbitrarily small probability of error as long as (36b) is satisfied.
- Having decoded its message, Rx_2 uses the DAF strategy to select the Kth conference message to Rx_1. The conference capacity allocated to this step is $R_{21}^{(K)} = \alpha C_{21}$.
- Having received the Kth conference message from Rx_2, Rx_1 can now decode its message using the information received at the first $K-1$ steps, and combining it with the information from the last step using the DAF decoding rule. This gives rise to (36a).

4) Combining All the Conference Rate Bounds: The bounds on $R_{12}'^{(k)}$, $k = 1,2,\ldots,K$, can be obtained as in Section IV-B6

$$
\begin{aligned}
C_{12} &= \sum_{k=1}^{K} R_{12}^{(k)}\\
&\ge I(\tilde{\mathbf{Y}}_1^{(K)}, \tilde{\mathbf{Y}}_2^{(K-1)}; Y_1|Y_2) + 2K\epsilon
\end{aligned}
$$

and similarly

$$
(1-\alpha)C_{21} \ge I(\tilde{\mathbf{Y}}_1^{(K)}, \tilde{\mathbf{Y}}_2^{(K-1)}; Y_2|Y_1) + 2K\epsilon
$$

where $(1-\alpha)C_{21}$ is the total capacity allocated to the first $K-1$ conference steps from Rx_2 to Rx_1. This provides the rate constraints on the conference auxiliary variables.

ACKNOWLEDGMENT

The authors would like to thank Gerhard Kramer and Shlomo Shamai for pointing out the conclusions of Appendix A and Section III-C. The authors would also like to thank the reviewers for their help in improving this paper.

REFERENCES

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

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

[3] P. Gupta and P. R. Kumar, "Toward an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[4] L. -L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[5] L. -L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.
[6] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. 41st Allerton Conf. Commun. Control Comput.*, Monticello, IL, 2003, pp. 1074–1083.
[7] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.
[8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
[9] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc. IEEE Int. Symp. Inf. Theory*, Lausanne, Switzerland, 2002, pp. 136–136.
[10] L. Lifeng, L. Ke, and H. El-Gamal, "The three-node wireless network: Achievable rates and cooperation strategies," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 805–828, Mar. 2006.
[11] C. T. K. Ng, I. Maric, A. J. Goldsmith, S. Shamai, and R. D. Yates, "Iterative and one-shot conferencing in relay channels," in *Proc. IEEE Inf. Theory Workshop*, Punta del Este, Uruguay, 2006, pp. 193–197.
[12] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 966–976, Aug. 2004.
[13] S. C. Draper, B. J. Frey, and F. R. Kschischang, "Interactive decoding of a broadcast message," in *Proc. 41st Allerton Conf.*, Urbana, IL, 2003.
[14] R. Dabora and S. D. Servetto, "Broadcast channels with cooperating decoders," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5438–5454, Dec. 2006.
[15] Y. Liang and V. V. Veeravalli, "Cooperative broadcast relay channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 900–928, Mar. 2007.
[16] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1522–1544, Apr. 2006.
[17] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
[18] A. El-Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.
[19] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2006.
[20] X. Bao and J. Li, "Decode-amplify-forward (DAF): a new class of forwarding strategy for wireless relay channels," in *Proc. 6th IEEE Workshop Signal Process. Adv. Wireless Commun.*, New York City, NY, 2005, pp. 816–820.
[21] K. S. Gomadam and S. A. Jafar, "On the capacity of memoryless relay networks," in *Proc. IEEE Int. Conf. Commun.*, Istanbul, Turkey, 2006, pp. 1580–1585.
[22] G. Kramer, "Distributed and layered codes for relaying," in *Proc. 39th Asilomar Conf. Signals Syst. Comput.*, 2005, pp. 1752–1756.
[23] J. N. Laneman and G. W. Wornell, "Energy-efficient antenna sharing and relaying for wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2000, vol. 1, pp. 7–12.
[24] I. Abou-Faycal and M. Medard, "Optimal uncoded regeneration for binary antipodal signalling," in *Proc. Commun. Theory Symp. IEEE Int. Conf. Commun.*, Paris, France, 2004, vol. 2, pp. 742–746.
[25] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.
[26] M. Motani, H.-F. Chong, and H. K. Garg, "Backward decoding strategies for the relay channel," in *Proc. MSRI Workshop: Math. Relaying Cooperat. Commun. Netw.*, Berkeley, CA, 2006.
[27] R. W. Yeung, *A First Course in Information Theory*. New York: Springer-Verlag, 2002.
[28] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[29] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
[30] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 441–445, May 1983.
[31] R. Dabora and S. D. Servetto, "On the rates for the general broadcast channel with partially cooperating receivers," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, 2005, pp. 2174–2178.
[32] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-44, no. 6, pp. 2524–2530, Oct. 1998.
[33] A. A. El-Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.
[34] A. H. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 735–740, Nov. 1985.
[35] R. Dabora and S. D. Servetto, "A multi-step conference for cooperative broadcast," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 2006, pp. 2190–2194.
[36] R. Dabora, "An information-theoretic study of cooperation in networks," Ph.D. dissertation, School Electr. Eng., Cornell Univ., Ithaca, NY, 2007.
[37] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 5, pp. 1106–1110, Sep. 1988.
[38] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2006.
[39] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inf. Control*, vol. 38, pp. 60–80, 1978.
[40] D. Chen and J. N. Laneman, "Modulation and demodulation for cooperative diversity in wireless systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1785–1794, Jul. 2006.
[41] M. Katz and S. Shamai, "Relaying protocols for two colocated users," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2329–2344, Jun. 2006.