

# Joint Source-Channel Cooperative Transmission over Relay-Broadcast Networks

Deniz Gündüz<sup>\*</sup>, Elza Erkip<sup>†</sup>, Andrea Goldsmith<sup>‡</sup>, H. Vincent Poor<sup>§</sup>

<sup>\*</sup>CTTC, Barcelona, Spain

<sup>†</sup>Dept. of Electrical Eng., Polytechnic Institute of New York University, Brooklyn, NY

<sup>‡</sup>Dept. of Electrical Eng., Stanford Univ., Stanford, CA

<sup>§</sup>Dept. of Electrical Eng., Princeton Univ., Princeton, NJ

Email: deniz.gunduz@cttc.es, elza@poly.edu, andrea@wsl.stanford.edu, poor@princeton.edu

## Abstract

Reliable transmission of a discrete memoryless source over a multiple-relay relay-broadcast network is considered. Motivated by sensor network applications, it is assumed that the relays and the destinations all have access to side information correlated with the underlying source signal. Joint source-channel cooperative transmission is studied in which the relays help the transmission of the source signal to the destinations by using both their overheard signals, as in the classical channel cooperation scenario, as well as the available correlated side information. Decode-and-forward (DF) based cooperative transmission is considered in a network of multiple relay terminals and two different achievability schemes are proposed: i) a regular encoding and sliding-window decoding scheme without explicit source binning at the encoder, and ii) a semi-regular encoding and backward decoding scheme with binning based on the side information statistics. It is shown that both of these schemes lead to the same source-channel code rate, which is shown to be the *source-channel capacity* in the case of i) a physically degraded relay network in which the side information signals are also degraded in the same order as the channel; and ii) a relay-broadcast network in which all the terminals want to reconstruct the source reliably, while at most one of them can act as a relay.

This work is supported in part by a Marie Curie grant funded by European Unions Seventh Framework Programme (FP7), by the Spanish Government under project TEC2010-17816 (JUNTOS), by the National Science Foundation under Grant CCF-0635177, by the DARPA ITMANET program under grant 1105741-1-TFIND and by the ARO under MURI award W911NF-05-1-0246.

DRAFT

## I. INTRODUCTION

A relay network consists of a source-destination pair and dedicated relay terminals that help the transmission of messages from the source to the destination. The classical relay channel model [1] focuses on the maximum channel coding rate that can be achieved with arbitrarily small probability of error. Most studies on the relay channel following [1] focus solely on the channel coding aspects of relaying, motivated by the improvement in the capacity, reliability or coverage extension provided by the relay terminals. On the other hand, in some applications, such as sensor networks, the relays might have partial information about the source signal obtained through their own sensing capabilities. This additional side information can be used to improve the end-to-end system performance. The advantages of exploiting the correlated side information in a sensor network from a purely source coding perspective have been illustrated in [2].

Here we consider the transmission of a discrete memoryless (DM) source over a DM relay-broadcast network with multiple relays and destinations, in which all the terminals in the network have access to their own correlated side information. The relays in the network are dedicated terminals whose only goal is to help the transmission of the source signal to the destinations. Some of the destinations also have channel inputs through which they can also relay the source signal to each other. The goal is the reliable (lossless) transmission of the underlying source signal to the destination(s), and the problem is to characterize the maximum number of source symbols per channel use<sup>1</sup> that can be transmitted reliably, called the *source-channel capacity*. This is a joint source-channel coding generalization of the classical relay network problem.

In this model the transmission scheme should exploit the availability of the side information at the network terminals as well as the overheard channel transmissions. Note that the classical channel cooperation ignores the side information at the terminals. However, this can lead to a significant performance loss. Consider, for example, a single relay channel in which there is no channel from the source terminal to the relay, i.e., the relay channel output is independent of the source terminal's channel input, while the relay has access to side information correlated with the underlying source signal. In this case the relay can still cooperate with the source terminal

<sup>1</sup>Here, the “channel use” refers to the use of the whole network, not the use of the separate source-relay or relay-destination channels. A more appropriate term would be the “network use”, but we stick here to the more common terminology in the literature.

by forwarding its side information to the destination. This special case is called the *one-helper problem*, and the benefits of cooperation in this setup are shown in [3] and [4]. Classical channel cooperation schemes cannot exploit the side information at the relay terminal, since their focus is on the processing of the overheard signal received by the relay.

Several channel coding techniques have been proposed for the relay channel [1]. In the decode-and-forward (DF) protocol the relays decode the underlying message, and cooperate with the source terminal to forward it to the destination. While not optimal in general, DF achieves the capacity in a physically degraded relay channel [1]. Here, we focus on the DF protocol in the joint source-channel transmission setting and propose multiple-relay extensions that exploit the side information at the relays and the destinations.

The DF protocols in the literature are categorized based on the codebook sizes and the decoding strategy. In *irregular encoding and successive decoding* [1], the relay and the source codebooks have different sizes and the destination applies successive decoding. In *regular encoding and sliding-window decoding*, introduced in [5], the source and the relay codebooks have the same size and the destination decodes each source message by using two consecutive channel blocks. Finally, in *regular encoding and backward decoding*, introduced in [6], the destination waits until all channel blocks are received, and decodes the messages starting from the last block and going backwards. For single source-single destination relay networks all encoding schemes lead to the same set of achievable rates despite having different delay behaviors; however, this is not the case in general, and backward decoding might lead to higher rates when there are multiple source terminals in the network [7]. DF channel coding is extended to multiple-relay networks in [8], [9], [10], [11] and [12]. While [9] and [12] consider irregular encoding, [8] and [10] study an extension of the regular encoding and sliding-window decoding scheme, and finally [11] and [13] extend the backward decoding strategy to multiple relays.

We propose two different *joint source-channel cooperation* protocols based on DF relaying. In particular we consider the joint source-channel coding extensions of the sliding-window [8], [10] and backward decoding schemes [11], [13]. These two transmission strategies differ in terms of the sizes of the codebooks used and the decoding delays, as well as the source encoding techniques; hence, these two schemes offer a tradeoff between decoding delay and complexity in the system, while achieving the same performance in terms of the source samples that can be transmitted per channel use, i.e., the source-channel code rate.

In transmitting a source signal to another user with correlated side information, the classical random coding approach developed by Slepian and Wolf in [14] is to use *binning* to reduce the amount of transmitted information. Source binning, which will be defined more rigorously in Section III, refers to dividing the possible source sequences into groups and sending only the index of the group rather than the index of the sequence. The decoder can then decode the source sequence using its side information together with the bin index. In the relay network setup with the DF protocol, due to the varying quality of side information at the terminals, separate binning is required for each terminal, and the corresponding bin indices are transmitted with channel codes at different rates. This will be the approach taken in the construction of the backward decoding scheme. However, we will see that transmission at the same rate is also possible without resorting to any binning operation. This will be the approach for building the sliding-window decoding scheme.

We should also remark that the proposed protocols are not expected to achieve the optimal performance in the general setting since our problem is a generalization of the classical relay network problem, which remains open. However, we prove that the proposed DF-based protocols achieve the optimal source-channel code rate, i.e., the source-channel capacity, in a physically degraded setting in which both the channel outputs and the side information sequences are degraded in the same order, and in a relay-broadcast network with one source and multiple destinations, such that at most one of the destinations can also act as a relay.

The problem of joint source-channel cooperative transmission has been previously studied for a single relay channel in [15], [16] and [17], and for a multiple-relay network in [18] and [19]. The techniques proposed in all these works are based on DF relaying with different transmission techniques. While semi-regular encoding and backward decoding with explicit binning at the source encoder is proposed in [15], irregular encoding/ successive decoding with and without explicit binning is considered in [16] and [17], respectively. In [18] a regular encoding/ sliding-window decoding scheme with explicit binning is considered in the multiple-relay setting. The lossy version of joint source-channel cooperative transmission is studied in [20].

The rest of the paper is organized as follows: In Section II, we introduce the system model and the problem. In Section III we illustrate the difference between binning and non-binning based joint source-channel coding schemes in a point-to-point scenario. The main results of the paper are stated in Section IV, in which the focus is on a relay network with a single destination



terminal. In Section V we extend the results to relay broadcast networks, and show that the proposed achievability technique based on joint source-channel DF scheme is optimal when there is only one destination with transmission capability. The paper is concluded in Section VI. The proofs of our main results are detailed in the Appendices.

In this paper we denote random variables by capital letters, sample values by the respective lower case letters, and alphabets by the respective calligraphic letters. The cardinality of set  $\mathcal{A}$  is denoted by  $|\mathcal{A}|$ . For  $k \leq n$ , the sequence  $(X_k, \dots, X_n)$  will be denoted by  $X_k^n$ , while  $X^n$  will be used for  $X_1^n$ . The complement of a certain element  $X_i$  in a vector  $X^n$  will be denoted by  $X_i^c \triangleq (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ .

Let  $\pi(\cdot)$  be an injective function<sup>2</sup> from a set  $\{1, \dots, N\}$  to set  $\{1, \dots, K\}$  with  $1 \leq N \leq K$ . For  $1 \leq i < j \leq N$ , we define  $\pi(i : j) \triangleq \{\pi(i), \pi(i+1), \dots, \pi(j)\}$ . We also define, for a set  $C = \{c_1, \dots, c_n\}$ ,  $n \in \mathbb{Z}^+$  and  $c_i \in \mathbb{Z}^+$ ,  $X_C \triangleq (X_{c_1}, \dots, X_{c_n})$ .

## II. PROBLEM SETUP

We first consider the relay network with a single source-destination pair and multiple relays. We have a network of  $K+2$  terminals (see Fig. 1): terminal  $T_0$  is the source terminal observing the source signal  $S_0$ , terminals  $T_i$  for  $i = 1, \dots, K$  are the  $K$  relay terminals each observing a different correlated side information signal  $S_i$ , and terminal  $T_{K+1}$  is the destination terminal with its own correlated side information signal  $S_{K+1}$ . The underlying DM relay channel is characterized by the conditional distribution

$$p(y_1^n, \dots, y_{K+1}^n | x_0^n, x_1^n, \dots, x_K^n) = \prod_{t=1}^n p_{Y_1, \dots, Y_{K+1} | X_0, \dots, X_K}(y_{1,t}, \dots, y_{K+1,t} | x_{0,t}, \dots, x_{K,t}),$$

where  $x_{i,t} \in \mathcal{X}_i$  and  $y_{i,t} \in \mathcal{Y}_i$ , respectively, are the channel input and output of terminal  $T_i$  at time  $t$ ; and the finite sets  $\mathcal{X}_i$  and  $\mathcal{Y}_i$  are the corresponding input and output alphabets. We denote this channel by  $(\mathcal{X}_0, \dots, \mathcal{X}_K, p(y_1^n, \dots, y_{K+1}^n | x_0^n, x_1^n, \dots, x_K^n), \mathcal{Y}_1, \dots, \mathcal{Y}_{K+1})$ .

We consider DM independent and identically distributed (i.i.d.) signals  $(S_0, \dots, S_{K+1})$  which are arbitrarily correlated according to a joint distribution  $p(s_0, \dots, s_{K+1})$  over a finite alphabet  $\mathcal{S}_0 \times \dots \times \mathcal{S}_{K+1}$ . The sequence  $\{S_{0,j}\}_{j=1}^\infty$  is denoted as the source sequence while  $\{S_{i,j}\}_{j=1}^\infty$ ,  $i = 1, \dots, K+1$ , is the side information sequence available at terminal  $T_i$ . We denote the set composed of the source and the side information signals by  $(\mathcal{S}_0, \dots, \mathcal{S}_{K+1}), p(s_0, \dots, s_{K+1})$ .

<sup>2</sup>A function  $f$  on a set  $\mathcal{A}$  is injective if for all  $a, b \in \mathcal{A}$ , if  $f(a) = f(b)$ , then  $a = b$ ; that is,  $f(a) = f(b)$  implies  $a = b$ .

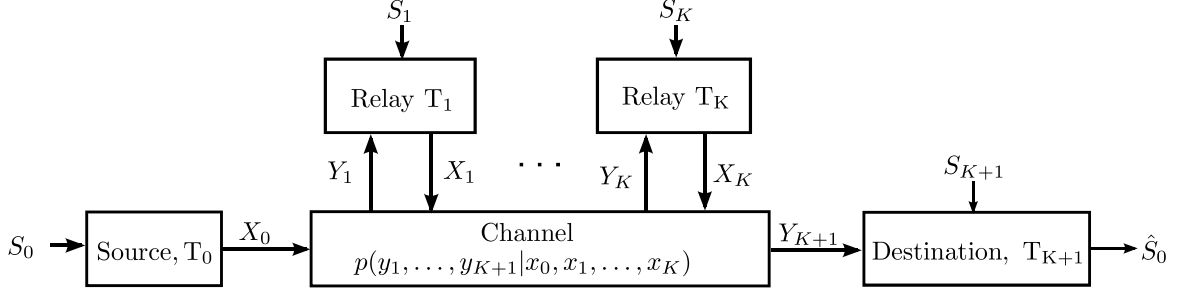


Fig. 1. Transmission of a discrete memoryless source over a relay network with correlated side information.

*Definition 1:* An  $(m, n)$  joint source-channel code with the sources  $(\mathcal{S}_0, \dots, \mathcal{S}_{K+1})$  and the relay network  $(\mathcal{X}_0, \dots, \mathcal{X}_K, p(y_1^n, \dots, y_{K+1}^n | x_0^n, x_1^n, \dots, x_K^n), \mathcal{Y}_1, \dots, \mathcal{Y}_{K+1})$  consists of

- 1) An encoding function  $f_0^{(m,n)} : \mathcal{S}_0^m \rightarrow \mathcal{X}_0^n$  at the source terminal  $T_0$  that maps its observation  $S_0^m$  to a channel codeword of length- $n$ , i.e.,  $X_0^n = f_0^{(m,n)}(S_0^m)$ .
- 2) A series of encoding functions for each relay terminal  $T_i, i = 1, \dots, K$ :  $f_i^{(m,n)} = \{f_{i,1}^{(m,n)}, \dots, f_{i,n}^{(m,n)}\}$ , such that the encoding function at time instant  $t$  depends on the previous channel outputs  $Y_i^{t-1}$  as well as its side information vector  $S_i^m$ . We have

$$X_{i,t} = f_{i,t}^{(m,n)}(Y_{i,1}, \dots, Y_{i,t-1}, S_i^m),$$

for  $1 \leq i \leq K$  and  $1 \leq t \leq n$ .

- 3) A decoding function at the destination terminal  $T_{K+1}$  which maps the channel output  $Y_{K+1}^n$  and its side information  $S_{K+1}^m$  to the estimate  $\hat{S}_0^m$  by

$$g^{(m,n)} : \mathcal{Y}_{K+1}^n \times \mathcal{S}_{K+1}^m \rightarrow \mathcal{S}_0^m, \quad (1)$$

$$\text{i.e., } \hat{S}_0^m = g^{(m,n)}(Y_{K+1}^n, S_{K+1}^m).$$

The goal of the network is to transmit the source message  $S_0^m$  to the destination terminal in a reliable manner. Reliability is based on the following definition of average probability of error.

*Definition 2:* The probability of error for an  $(m, n)$  code is defined as

$$P_e^{(m,n)} = \Pr \left\{ \hat{S}_{K+1}^m \neq S_0^m \right\},$$

where the averaging is over both the source and the channel distributions.

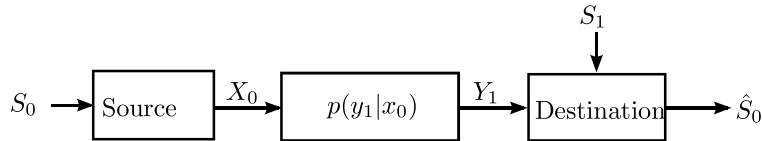


Fig. 2. Transmission of a discrete memoryless source over a point-to-point channel with correlated side information.

*Definition 3:* We say that the source-channel code rate of  $r$  source samples per channel use is *achievable* if there exists a sequence of  $(m, n)$  codes satisfying  $\frac{m}{n} \geq r$  whose probability of error vanishes with increasing block size, i.e.,  $P_e^{(m,n)} \rightarrow 0$  as  $m, n \rightarrow \infty$ .

In [21] we have defined the *source-channel rate* as the number of channel uses required to transmit each source sample, that is, as the inverse of the *source-channel code rate* we have defined here. The latter definition is in accordance with the definition of the rate of a joint source-channel code used in [22]. Similar to the definition of channel capacity, we define a source-channel capacity of a network which considers both the source and the channel characteristics.

*Definition 4:* The *source-channel capacity* of a network is defined as the supremum of all achievable source-channel code rates.

### III. POINT-TO-POINT CHANNEL: TO BIN OR NOT TO BIN

In this section we focus on the point-to-point channel setup to introduce the basic concepts such as binning and source-channel separation, which will be helpful in understanding the coding schemes proposed for the relay network. The point-to-point channel without side information at the receiver was studied by Shannon, who proved the optimality of source-channel separation in this setup [23]. The point-to-point channel with side information at the receiver was studied by Shamai and Verdú in [24]. In the system model introduced in Section II, the point-to-point channel with receiver side information corresponds to the case with  $K = 0$  (see Fig. 2). It is shown in [24] that a source-channel code rate  $r$  is achievable if there exists an input distribution  $p(x_0)$  such that  $r < \frac{I(X_0; Y_1)}{H(S_0|S_1)}$ ; and conversely, if the source-channel code rate  $r$  is achievable then there exists an input distribution  $p(x_0)$  such that  $r \leq \frac{I(X_0; Y_1)}{H(S_0|S_1)}$ . Note that  $I(X_0; Y_1)$  is maximized by the capacity achieving input distribution. Hence, equivalently, the source-channel capacity of this system is  $\frac{C}{H(S_0|S_1)}$ , where  $C$  is the channel capacity. Moreover, it is not very difficult to see that the source-channel separation theorem holds for this setup, that is, any source-channel

code rate less than the source-channel capacity,  $\frac{C}{H(S_0|S_1)}$  can be achieved by first applying source coding, and then transmitting the compressed source bits over the channel using a capacity achieving channel code. As pointed out in [24], the source encoder in the case of separation is a Slepian-Wolf encoder that operates at the conditional entropy rate of  $H(S_0|S_1)$  rather than the source entropy rate of  $H(S_0)$ , due to the availability of the correlated side information at the receiver.

Slepian-Wolf compression involves *binning* of the source outcomes. The source encoder randomly distributes all possible source output sequences  $S_0^m$  into  $2^{mH(S_0|S_1)}$  bins, that is, it independently assigns an index uniformly distributed over  $\{1, \dots, 2^{mH(S_0|S_1)}\}$  to each of the possible source output sequences. The bin index for each source sequence is transmitted over the channel by assigning a channel codeword to each bin index. Having decoded the channel codeword, and hence, the bin index, correctly, the receiver outputs the source outcome in the corresponding bin that is jointly typical with its side information sequence.

We next present a coding scheme that generalizes this separate source-channel coding approach. In this generalized scheme, we randomly distribute all  $S_0^m$  sequences into  $M = 2^{mR}$  bins, where  $R$  is not necessarily equal to  $H(S_0|S_1)$ . Let  $\mathcal{B}(i)$  be the set of sequences allocated to bin  $i$ . Then, we generate  $M$  independent channel codewords of length  $n$  according to distribution  $\prod_{t=1}^m p(x_{0,t})$ , and enumerate these codewords as  $x_0^n(w)$  for  $w = 1, \dots, M$ . This constitutes the only codebook in the system. Encoding is done as in the separation scheme. The transmitter finds the index  $i$  of the bin to which  $s_0^m$  belongs, and transmits over the channel the codeword  $x_0^n(i)$ .

In the classical source-channel separation approach, the channel decoder, upon observing the channel output, decides a single codeword index, and conveys this index to the source decoder. This index corresponds to the bin index to which the source sequence belongs. Then the source decoder estimates the source sequence using the bin index and the side information sequence. For the channel transmission to be successful with high probability, the rate of transmission should be less than the channel capacity. Hence, in the separation approach, it is advantageous to reduce the number of possible indices to be transmitted through source binning to achieve higher source-channel code rates. However, in the generalized scheme, we consider a joint source-channel decoder, following the approach in [25]. The decoder at the receiver looks for an index  $i$  for which  $x_0^n(i)$  and  $Y_1^n$  are jointly typical, and at the same time, there exists exactly one

source sequence  $\hat{s}^m$  in bin  $i$  that is jointly typical with  $S_1^m$ . This source sequence is the output of the decoder. Note that, due to the joint nature of decoding using both the side information and the channel output, this is not separate source and channel coding in the strict sense. We refer the readers to [25] and [21] for more discussion on source-channel separation in multi-terminal scenarios.

We have an error if there exists no or more than one such bin index  $i$ , or if there exists more than one jointly typical sequence within bin  $i$ . The probability that there is no bin index satisfying the joint typicality condition vanishes as  $n$  grows. The probability of having no jointly typical source sequence within the correct bin also vanishes since  $S_0^m$  and  $S_1^m$  are jointly typical with high probability as  $m$  grows. The probability of having another jointly typical source sequence in the same bin as  $S_0^m$  is bounded by

$$|\mathcal{B}(i) \cap A_\epsilon^m(S_0)| 2^{-m(I(S_0;S_1)-3\epsilon)} \leq 2^{m(H(S_0)+\epsilon)} 2^{-mR} 2^{-m(I(S_0;S_1)-3\epsilon)}, \quad (2)$$

in which  $A_\epsilon^m(S_0)$  denotes the set of  $\epsilon$ -typical  $n$ -tuples according to  $P_{S_0}$ . Using the classical arguments on typical sets [26], it is possible to show that (2) goes to zero if  $R \geq H(S_0|S_1)$ .

We also have an error if there exists another bin index  $j$  satisfying the joint typicality conditions. The probability of this event can be bounded by

$$2^{mR} 2^{-n(I(X_0;Y_1)-3\epsilon)} |\mathcal{B}(i) \cap A_\epsilon^m(S_0)| 2^{-m(I(S_0;S_1)-3\epsilon)} \leq 2^{-n(I(X_0;Y_1)-3\epsilon)} 2^{m(H(S_0|S_1)-2\epsilon)},$$

which goes to zero if  $mH(S_0|S_1) < nI(X_0;Y_1)$ . Hence, any rate  $r$  satisfying  $r < \frac{I(X_0;Y_1)}{H(S_0|S_1)}$  is achievable.

Now, we have a set of coding schemes each with a different number of source bins, that is, with different  $R$  values satisfying  $R \geq H(S_0|S_1)$ . As suggested in [27], the “joint” decoding operation considered in the generalized scheme can equivalently be viewed as a separate source and channel decoding scheme, in which the channel decoder is a list decoder, which outputs the list of bin indices  $i$  for which  $x_0^n(i)$  and  $Y_1^n$  are jointly typical. This list decoding approach includes separate source-channel coding as a special case with  $R = H(S_0|S_1)$ ; in which case we have, with high probability, a single element in the list, i.e., there exists only a single bin index whose channel input codeword is typical with the channel output.

We want to point out here that, on the other extreme, this generalized scheme works without any binning, such that we generate an independent channel codeword for each possible source

outcome, i.e.,  $R = \log |\mathcal{S}_0|$ . From a practical point of view, this can be seen as transferring the complexity of binning from the encoder to the decoder, which now needs to apply joint decoding or list decoding. From a theoretical point of view, since the decoder only outputs typical source sequences as its estimate, there is no point in having more than  $2^{m(H(S_0)+\epsilon)}$  bins as, otherwise, there would be bins without any typical source sequences, i.e., we have  $R \leq H(S_0)$ . Hence, in the rest of the paper, schemes with  $R = H(S_0)$  are considered as *no-binning* schemes.

In the case of a point-to-point channel, the only difference between separate source and channel coding with binning, and joint decoding with no-binning is the operation at the encoder and the decoder. However, as we will see in the following sections, in the case of relay networks these two approaches require different transmission protocols, and hence, obtain different delay performances even though they still achieve the same source-channel code rate performance.

#### IV. MAIN RESULTS

In this section we address the main problem of source transmission over relay networks. We state our results for the single destination setup in this section while the proofs are given in the following sections. The first theorem provides an achievability result.

*Theorem 1:* For the DM relay network with correlated relay and destination side information, the source-channel code rate  $r$  is achievable if, for all  $i = 1, \dots, N$ ,

$$r < \frac{I(X_{\pi(0:i-1)}; Y_{\pi(i)} | X_{\pi(i:N-1)})}{H(S_0 | S_{\pi(i)})}, \quad (3)$$

for an  $N$  satisfying  $1 \leq N \leq K+1$ , and for some injection  $\pi(\cdot)$  from  $\{0, \dots, N\}$  to  $\{0, \dots, K+1\}$  such that  $\pi(0) = 0$ ,  $\pi(N) = K+1$ , and for some input distribution  $p(x_0, \dots, x_K)$  and  $\mathcal{X}_{K+1}$  is a constant.

Note in (3) that we impose decoding constraints for only a subset of the relays in the network together with the destination, such that only  $N-1$  relays decode rather than all  $K$  of them. Excluding some of the relays from cooperation, the relays with a poor channel from the source terminal as well as poor side information, can potentially increase the rate in the case of DF relaying, as these relays can decode only at very low rates.

In the appendix we provide two different proofs for the achievability of Theorem 1. Both proofs are based on DF relaying in the joint source-channel setting, that is, the source vector  $S_0^m$  is decoded in a lossless fashion by all the terminals participating in cooperation. The first

proof is based on block-Markov regular encoding and sliding-window decoding without explicit binning. This achievable scheme is based on the joint source-channel decoding scheme with no binning introduced in Section III. Here, we combine this technique with block Markov encoding in a multiple-relay setup. The typical source outcomes are mapped directly to different channel codewords rather than binning the source outputs prior to channel coding<sup>3</sup>. The decoders at each relay apply joint source-channel decoding (or, separate channel and source decoders in which the channel decoder outputs a list of possible inputs rather than a single codeword index). Each relay finds the unique index for which the corresponding source codeword is jointly typical with its side information while the corresponding channel codewords are jointly typical with the received channel vectors in the preceding blocks. This is a regular coding scheme since all the terminals in the network use a codebook of the same size, which is equal to the number of typical source outputs. The details of this achievability proof are given in Appendix A.

The second coding scheme, which was studied in [15] for a single relay channel, uses explicit binning at the source encoder and channel codes of different sizes for each terminal in the network. We call this scheme semi-regular encoding with backward decoding. The source is compressed (by binning) for each separate side information signal in the network, and hence a different rate of information is transmitted to each user; however, the rate of the channel codes for the terminals that have already decoded the message and are cooperating to forward it to the next terminal are the same. This is why we call this coding scheme a semi-regular encoding scheme. For decoding we use nested backward decoding [11], [13]. The detailed analysis of this achievability scheme is given in Appendix B.

These two coding schemes essentially differ in terms of the delay in decoding, while each message block is decoded after a delay of  $K$  channel blocks in the case of sliding-window decoding, the delay is much larger in the case of backward decoding, since the destination can start decoding only after receiving all the channel blocks. In the case of pure channel coding, the two schemes have exactly the same encoding structure; hence, the tradeoff is between the delay and the complexity. However; in the joint source-channel coding setting, there is another difference between the proposed coding techniques. While backward decoding works together

<sup>3</sup>This is equivalent to binning all source outcomes into  $R = 2^{mH(S_0)}$  bins in the scheme introduced in Section III, so that there is one typical source outcome in each bin.

with explicit binning, sliding-window decoding is based on matching the source outcome directly to a channel input. Therefore, in practical systems, the backward decoding scheme can be directly implemented using the existing point-to-point source codes and DF channel codes for the relay channel. However, the sliding-window coding scheme requires building new codes that implement the joint source-channel encoding and decoding techniques in practice.

In the following theorem it is shown that the proposed schemes achieve the source-channel capacity in a physically degraded relay network with degraded side information sequences. The definition of a physically degraded relay network is given below.

*Definition 5:* A discrete memoryless relay network is said to be *physically degraded* if

$$p(y_{i+1}, \dots, y_{K+1} | y_i, x_0, \dots, x_K) = p(y_{i+1}, \dots, y_{K+1} | y_i, x_i, \dots, x_K) \quad (4)$$

for all  $i = 1, \dots, K$ , or equivalently if

$$(X_0, \dots, X_{i-1}) \rightarrow (Y_i, X_i, \dots, X_K) \rightarrow (Y_{i+1}, \dots, Y_{K+1})$$

forms a Markov chain for all  $i = 1, \dots, K$ .

*Theorem 2:* For a physically degraded relay network in which the side information sequences also form a Markov chain in the same order, i.e.,

$$S_0 \rightarrow S_1 \rightarrow \dots \rightarrow S_{K+1},$$

the source-channel capacity is given by

$$\sup_{p(x_0, x_1, \dots, x_K)} \min_{i=1, \dots, K+1} \frac{I(X_0^{i-1}; Y_i | X_i^{K+1})}{H(S_0 | S_i)}. \quad (5)$$

*Proof:* The converse for degraded relay networks follow from the cut-set bound. Consider the set  $\mathcal{S}_i = \{T_0, \dots, T_{i-1}\}$  and assume that the terminals in  $\mathcal{S}_i$  all have access to the source vector  $S_0^m$ ; hence, they can cooperate perfectly for transmitting  $S_0^m$ . We further assume that the remaining terminals can also cooperate perfectly by pooling all the available side information vectors  $S_i^m, \dots, S_{K+1}^m$  as well as their received channel outputs  $Y_i^m, \dots, Y_{K+1}^m$ . This reduces to a point-to-point scenario for which the following is a necessary condition for reliable transmission:

$$H(S_0 | S_i, \dots, S_{K+1}) \leq r I(X_0^{i-1}; Y_i^{K+1} | X_i^{K+1}).$$



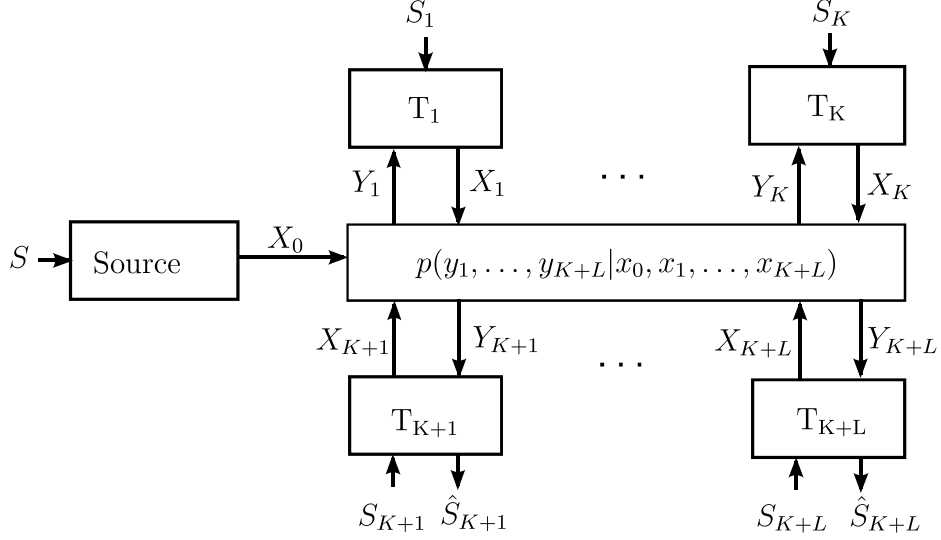


Fig. 3. Transmission of a discrete memoryless source over a relay-broadcast network with  $K$  relay terminals and  $L$  destination terminals.

From the degradedness assumption of the side information vectors we have

$$H(S_0 | S_i, \dots, S_{K+1}) = H(S_0 | S_i),$$

and from the physically degraded channel assumption we have

$$I(X_0^{i-1}; Y_i^{K+1} | X_i^{K+1}) = I(X_0^{i-1}; Y_i | X_i^{K+1}).$$

We complete the proof of the theorem by considering all  $S_i$  sets corresponding to  $i = 1, \dots, K+1$ . ■

## V. EXTENSION TO RELAY-BROADCAST NETWORKS

It is possible to generalize the achievability results in Section IV to a relay-broadcast network, in which there are multiple receivers interested in decoding the source samples in a lossless fashion (see Figure 3). In this setting, terminals  $T_{K+1}, \dots, T_{K+L}$  not only decode the source sequence, but can also relay the decoded sequence for each other. We call this more general network a *relay-broadcast network*. The following theorem provides an achievable source-channel code rate for a relay-broadcast network. The proof follows similarly to the proof of Theorem 1, in which a subset of the relays and the destination terminals apply the DF protocol in some given order.

In this model, the DM channel is characterized by the conditional distribution

$$p(y_1^n, \dots, y_{K+L}^n | x_0^n, x_1^n, \dots, x_{K+L}^n).$$

The source and side information sequences are DM and characterized by the joint distribution

$$p(s_0, \dots, s_{K+1})$$

over a finite alphabet  $\mathcal{S}_0 \times \dots \times \mathcal{S}_{K+L}$ . Terminal  $T_0$  maps its observation  $S_0^m$  to a channel codeword of length- $n$  by the encoding function  $f_0^{(m,n)} : \mathcal{S}^m \rightarrow \mathcal{X}_0^n$ . The terminal  $T_i$  has encoding functions  $f_i^{(m,n)} = \{f_{i,1}^{(m,n)}, \dots, f_{i,n}^{(m,n)}\}$  such that  $X_{i,t} = f_{i,t}^{(m,n)}(Y_{i,1}, \dots, Y_{i,t-1}, S_i^m)$ , for  $1 \leq i \leq K+L$  and  $1 \leq t \leq n$ .

The decoder at a destination terminal  $T_j$ ,  $j = K+1, \dots, K+L$ , maps the channel output  $Y_j^n$  and its side information  $S_j^m$  to an estimate  $\hat{S}_j^m$  by the decoding function  $\hat{S}_j^m = g_j^{(m,n)}(Y_j^n, S_j^m)$ . The probability of error is defined as  $P_e^{(m,n)} = \Pr \left\{ \bigcup_{j=K+1}^{K+L} \{\hat{S}_j^m \neq S_0^m\} \right\}$ . The definition of an achievable source-channel code rate is similar to Definition 4 using this new probability of error definition.

*Theorem 3:* For the DM relay-broadcast network with correlated relay and destination side information as in Figure 3, the source-channel code rate  $r$  is achievable if, for all  $i = 1, \dots, N$ ,

$$r < \frac{I(X_{\pi(0:i-1)}; Y_{\pi(i)} | X_{\pi(i)}, \dots, X_{\pi(K+L)})}{H(S_0 | S_{\pi(i)})}, \quad (6)$$

for an  $N$  satisfying  $L \leq N \leq K+L$ , and for some injection  $\pi(\cdot)$  from  $\{0, \dots, N\}$  to  $\{0, \dots, K+L\}$  such that  $\pi(0) = 0$  and  $\{K+1, K+2, \dots, K+L\} \subseteq \pi(1:N)$ , and some input distribution  $p(x_0, \dots, x_{K+L})$ .

When all the terminals in the network wish to decode the source sequence, i.e.,  $K = 0$  in the relay-broadcast network, and the channel inputs of these terminals are deterministic, i.e.,  $|\mathcal{X}_{K+1}| = \dots = |\mathcal{X}_{K+L}| = 1$ , then the problem reduces to the problem of broadcasting a common source to multiple terminals each with different side information. This problem is studied in [25] and it is shown that the source-channel capacity in this setup is given by

$$\sup_{p(x_0)} \min_{i=1, \dots, L} \frac{I(X_0; Y_i)}{H(S_0 | S_i)}. \quad (7)$$

For this special case, our achievable scheme based on regular encoding and sliding-window decoding reduces to the coding scheme in [25], whereas the semi-regular encoding and backward

decoding scheme, introduced in the proof of Theorem 1 provides an alternative optimal coding scheme for the broadcasting problem.

Another special case for which the source-channel capacity can be completely characterized is when all the terminals are interested in receiving the source sequence and there is only a single terminal with transmission capability, that is,  $K = 0$  and  $|\mathcal{X}_2| = \dots = |\mathcal{X}_L| = 1$ . We have the following result for this setup.

*Lemma 1:* For the DM single-relay broadcast network with correlated relay and destination side information at the terminals, i.e.,  $K = 0$ ,  $L \geq 2$  and  $|\mathcal{X}_2| = \dots = |\mathcal{X}_L| = 1$ , the source-channel capacity is given by

$$\sup_{p(x_0, x_1)} \min \left\{ \frac{I(X_0; Y_1 | X_1)}{H(S_0 | S_1)}, \frac{I(X_0, X_1; Y_2)}{H(S_0 | S_2)}, \dots, \frac{I(X_0, X_1; Y_L)}{H(S_0 | S_L)} \right\}. \quad (8)$$

*Proof:* While the achievability is a direct result of Theorem 3, the converse follows from the cut-set bound. The first term in the minimization comes from the cut around terminal  $T_1$ , and the following terms follow from the cuts around each of the terminals  $T_2, \dots, T_L$ . ■

## VI. CONCLUSION

We have considered the reliable transmission of a discrete memoryless source signal over a cooperative multiple-relay relay-broadcast network in which the relays and the destinations all have access to a different side information signal correlated with the source signal. We have defined the source-channel code rate of a code for such a system as the number of source symbols per channel use that can be transmitted reliably to the destination. The goal is to identify the maximum source-channel code rate, which we have defined as the source-channel capacity.

We have developed two different joint source-channel cooperation schemes which generalize decode-and-forward relaying to joint source-channel cooperation in multiple-relay networks, and we have identified the achievable source-channel code rate for these schemes. The first scheme does not use explicit source binning and is based on regular block-Markov encoding with sliding-window joint source-channel decoding, while the second scheme applies explicit source binning and uses separate source and channel decoders based on semi-regular block-Markov encoding and backward decoding. Our schemes illustrate that these two encoding schemes, apart from leading to different amounts of delay in decoding, also require two different types of implementation

in the joint source-channel coding context. The consequences of this variation on the design of the practical codes is the subject of future research. Moreover, we have proven the optimality of DF relaying in the joint source-channel setting for a physically degraded relay channel with degraded side information, and for a relay broadcast network when all the terminals are interested in decoding the source signal, but at most one of the destinations has the transmission capability.

## APPENDIX A

### PROOF OF THEOREM 1: REGULAR ENCODING AND SLIDING-WINDOW DECODING

For a given  $(n, m)$  pair, we build the code as follows. Consider  $B - K \geq 0$  source blocks, each consisting of  $m$  source samples, enumerated as  $S_0^m(b)$ ,  $b = 1, \dots, B - K$ . These source blocks will be transmitted to the destination  $T_{K+1}$  over  $B$  channel blocks, each consisting of  $n$  channel uses. This corresponds to a source-channel code rate of  $Bn/(B - K)m$  channel uses per source sample. Note that this rate gets arbitrarily close to  $n/m$  as  $B \rightarrow \infty$ .

Without loss of generality, we consider the special injection  $\pi(i) = i$  for  $i = 0, \dots, K + 1$ , and present the achievable scheme for this permutation. Generalization to any other subset of the relays and any other decoding order follows similarly.

Fix  $p(x_0, \dots, x_K)$  such that (3) holds. We use superposition block Markov encoding and sliding-window decoding.

*Source code generation:* Generate at random  $M = 2^{m(H(S_0) + \epsilon)}$  i.i.d. source codewords  $s^m(w_0)$  in  $\mathcal{S}_0^m$ ,  $w_0 \in [1, M]$ , each drawn according to the distribution  $\prod_{t=1}^m p(s_{0,t})$ . This constitutes the source codebook.

*Channel code generation:* Generate at random  $M$  i.i.d. channel codewords  $x_{K-1}^n(w_{K-1})$  in  $\mathcal{X}_{K-1}^n$ ,  $w_{K-1} \in [1, M]$ , each drawn according to the distribution  $\prod_{t=1}^n p(x_{K-1,t})$ . This constitutes the random channel codebook of relay  $T_{K-1}$ .

Then for each  $x_{K-1}^n(w_{K-1})$ , generate at random  $M$  conditionally i.i.d. channel codewords  $x_{K-2}^n(w_{K-2}|w_{K-1})$ ,  $w_{K-2} \in [1, M]$ , each drawn according to the distribution  $\prod_{t=1}^n p(x_{K-2,t}|x_{K-1,t}(w_{K-1}))$ . This constitutes the random channel codebook of relay  $T_{K-2}$ .

We continue the generation of codebooks sequentially for the terminals  $T_{K-3}, T_{K-4}, \dots, T_0$ . For each tuple of

$$\{x_{i+1}^n(w_{i+1}|w_{i+2}, \dots, w_{K-1}), x_{i+2}^n(w_{i+2}|w_{i+3}, \dots, w_{K-1}), \dots, x_{K-1}^n(w_{K-1})\}$$

Terminal	Block 1	Block 2	Block 3
$T_0$	$x_0^n(w_0(1) 1, 1)$	$x_0^n(w_0(2) w_0(1), 1)$	$x_0^n(w_0(3) w_0(2), w_0(1))$
$T_1$	$x_1^n(1 1)$	$x_1^n(\hat{w}_0^1(1) 1)$	$x_1^n(\hat{w}_0^1(2) \hat{w}_0^1(1))$
$T_2$	$x_2^n(1)$	$x_2^n(1)$	$x_2^n(\hat{w}_0^2(1))$
Terminal	Block $B-2$	Block $B-1$	Block $B$
$T_0$	$x_0^n(w_0(B-2) w_0(B-3), w_0(B-4))$	$x_0^n(1 w_0(B-2), w_0(B-3))$	$x_0^n(1 1, w_0(B-2))$
$T_1$	$x_1^n(\hat{w}_0^1(B-3) \hat{w}_0^1(B-4))$	$x_1^n(\hat{w}_0^1(B-2) \hat{w}_0^1(B-3))$	$x_1^n(1 \hat{w}_0^1(B-2))$
$T_2$	$x_2^n(\hat{w}_0^2(B-4))$	$x_2^n(\hat{w}_0^2(B-3))$	$x_2^n(\hat{w}_0^2(B-2))$

Fig. 4. Channel codeword assignment for the regular encoding and sliding-window decoding scheme with  $K = 2$  relays. We transmit  $B - 2$  source blocks to the destination in  $B$  channel blocks.

generate  $M$  conditionally independent channel codewords  $x_i^n(w_i|w_{i+1}, \dots, w_{K-1})$ ,  $w_i \in [1, M]$ , each drawn according to the distribution  $\prod_{t=1}^n p(x_{i,t}|x_{i+1,t}(w_{i+1}, \dots, w_{K-1}), \dots, x_{K-1,t}(w_{K-1}))$ . This constitutes the random channel codebook of terminal  $T_i$ , for  $i = K - 3, K - 4, \dots, 0$ .

Finally, the channel codebook generation process is repeated independently  $K - 1$  times, and these codebooks are used sequentially over different channel blocks, so that when the joint typicality decoding is applied simultaneously over  $K$  consecutive channel blocks, the decoding errors corresponding to different blocks are independent of each other.

*Encoding:* At channel block  $b$ , for  $b = 1, \dots, B - K$ , the source terminal  $T_0$  finds the index  $w_0(b)$  of the source outcome  $s_0^m(b)$ . The index  $w_0(b)$  is set to 1 if the source realization is not typical. We set  $w_0(b) = 1$  for  $b > B - K$  for notational convenience.

From the decoding procedure, which will be presented next, at the beginning of block  $b$  for  $b = 1, \dots, B$ , terminal  $T_i$ ,  $i = 0, \dots, K$ , has the estimates  $\hat{w}_i(b - k + 1)$  of  $w_0(b - k + 1)$  for  $k \geq i + 1$ , where we let  $\hat{w}_0(i) = w_0(i)$  for  $i = 1, \dots, b$ . Terminal  $T_i$  then sends  $x_i^n(\hat{w}_i(b - i)|\hat{w}_i(b - i - 1), \dots, \hat{w}_i(b - K))$  over channel block  $b$  using the codebook whose turn has come, where we set  $\hat{w}_i(b') = w_0(b') = 1$  for every  $b' < 1$  and  $b' > b - K + 1$ . See Fig. 4 for an illustration of the encoding procedure in a network with  $K = 2$  relays.

*Decoding:* At the end of block  $b$ , for  $b = 1, \dots, B$ , terminal  $T_i$ ,  $i = 1, \dots, K + 1$ , declares  $\hat{w}_i(b - i + 1) = w$  if there exists a unique index  $w \in [1, M]$  for which

$$(s^m(w), S_i^m(b - i + 1)) \in A_e^m(S_0, S_i)$$

and

$$(x_{i-1-j}^n(w|\hat{w}_i(b - i), \dots, \hat{w}_i(b - j - K)), \dots, x_{K-1}^n(\hat{w}_i(b - j - K)), Y_i^n(b - j)) \in A_e^n(X_{i-1-j}, \dots, X_K, Y_i),$$

in all the blocks  $b - j$  for  $j = 0, \dots, i - 1$ . An error is declared if no or more than one such index is found. Note that terminal  $T_i$ ,  $i = 1, \dots, K + 1$ , attempts decoding at channel blocks  $i \leq b \leq B - K + i - 1$ .

*Analysis of probability of error:* The probability of not having a typical source outcome vanishes as  $m$  increases. Hence, in the rest of the error analysis we will focus on the analysis of error for decoding the typical source realization at the nodes.

Denote by  $P_i(b)$  the probability of the event that a decoding error is made at terminal  $T_i$ ,  $i = 1, \dots, K + 1$ , in block  $b$ ,  $b = 1, \dots, B$ , conditioned on the event that no decoding error is made in the previous blocks. That is, we have

$$P_i(b) \triangleq \Pr\{\hat{w}_i(b - i + 1) \neq w_0(b - i + 1) | E^c(b - 1)\},$$

where

$$E^c(b) \triangleq \{\hat{w}_i(b' - i + 1) = w_0(b' - i + 1) \text{ for all } b' = 1, \dots, b \text{ and } i = 1, \dots, K + 1\}.$$

Then the probability of error  $P_e$  can be bounded as follows

$$P_e = \sum_{b=1}^B \Pr\{\hat{w}_i(b - i + 1) \neq w_0(b - i + 1) \text{ for some } i \in \{1, \dots, K\} | E^c(b - 1)\} \cdot \Pr\{E^c(b - 1)\}, \quad (9)$$

$$\leq \sum_{b=1}^B \sum_{i=1}^K P_i(b) \cdot \Pr\{E^c(b - 1)\}. \quad (10)$$

Assuming no decoding error is made in the previous blocks, to calculate  $P_i(b)$  we can assume

$$\hat{w}_i(b' - i + 1) = w_0(b' - i + 1)$$

for all  $b' = 1, \dots, b - 1$ . Then terminal  $T_i$ ,  $i = 1, \dots, K$ , declares  $\hat{w}_i(b - i + 1) = w$ , for  $b - i + 1 \geq 1$ , if  $w$  is the unique index  $w \in [1, M]$  such that

$$(s^m(w), S_i^m(b - i + 1)) \in A_\epsilon^m(S_0, S_i) \quad (11)$$

and

$$\begin{aligned} (x_{i-1-j}^n(w|w(b-i), \dots, w(b-j-K)), \dots, x_{K-1}^n(w(b-j-K)), Y_i^n(b-j)) \\ \in A_\epsilon^n(X_{i-1-j}, \dots, X_K, Y_i) \end{aligned} \quad (12)$$

hold simultaneously for all the blocks  $b - j$  for  $j = 0, 1, \dots, i - 1$ .

We define the following sets

$$E_i^0(b) \triangleq \{w \in [1, M] : w \text{ satisfies (11)}\}, \quad (13)$$

$$E_{i,j}(b) \triangleq \{w \in [1, M] : w \text{ satisfies (12)}\}, \quad (14)$$

$$E_i(b) \triangleq \bigcap_{j=0}^{i-1} E_{i,j}(b), \quad (15)$$

for  $i = 1, \dots, K$  and  $b = 1, \dots, B$ . Then,  $P_i(b)$  can be written as

$$P_i(b) = \Pr \{w_0(b - i + 1) \notin E_i^0(b) \text{ or } w_0(b - i + 1) \notin E_i(b) \\ \text{or for some } w' \in E_i^0(b) \cap E_i(b) \text{ but } s^m(w') \neq S_0^m(b - i + 1) | E^c(b - 1)\} \quad (16)$$

$$\leq \frac{1}{\Pr\{E^c(b - 1)\}} [\Pr\{w_0(b - i + 1) \notin E_i^0(b)\} + \Pr\{w_0(b - i + 1) \notin E_i(b)\} \\ + \Pr\{\text{for some } w' \in E_i^0(b) \cap E_i(b) \text{ but } s^m(w') \neq s_0^m(b - i + 1)\}] . \quad (17)$$

Hence, the probability of error can be bounded as

$$P_e \leq \sum_{b=1}^B \sum_{i=1}^K [\Pr\{w_0(b - i + 1) \notin E_i^0(b)\} + \Pr\{w_0(b - i + 1) \notin E_i(b)\} \\ + \Pr\{\text{for some } w' \in E_i^0(b) \cap E_i(b) \text{ but } s^m(w') \neq s_0^m(b - i + 1)\}] . \quad (18)$$

The first two arguments in the above summation can be made arbitrarily small for large enough  $m$  and  $n$  [26]. On the other hand, we have

$$\Pr\{\text{for some } w' \in E_i^0(b) \cap E_i(b) \text{ but } s^m(w') \neq s_0^m(b - i + 1)\} \\ \leq \sum_{w' \in [1, M], w' \neq w_0(b - i + 1)} \Pr\{w' \in E_i^0(b) \cap E_i(b)\} \quad (19)$$

$$= \sum_{w' \in [1, M], w' \neq w_0(b - i + 1)} \Pr\{w' \in E_i^0(b)\} \cdot \Pr\{w' \in E_i(b)\} \quad (20)$$

$$= \sum_{w' \in [1, M], w' \neq w_0(b - i + 1)} \Pr\{w' \in E_i^0(b)\} \cdot \prod_{j=0}^{i-1} \Pr\{w' \in E_{i,j}(b)\}, \quad (21)$$

$$\leq (M - 1) 2^{-m(I(S_0; S_i) - 3\epsilon)} 2^{-n(I(X_0, \dots, X_{i-1}; Y_i | X_i, \dots, X_{K-1}) - 6i\epsilon)} \quad (22)$$

where (20) follows from the independence among the source and the channel codebooks; (21) follows from the independence among the channel codebooks at consecutive channel blocks; and (22) follows from the chain rule of mutual information and the following inequality:

$$\Pr\{w' \in E_{i,j}(b)\} \leq 2^{-n(I(X_{i-1-j}; Y_i | X_{i-j}, \dots, X_{K-1}) - 6\epsilon)}$$

for  $w' \neq w_0(b - i + 1)$ . Finally, substituting the value of  $M$ , we get

$$\begin{aligned} \Pr\{\text{for some } w' \in E_i^0(b) \cap E_i(b) \text{ but } s^m(w') \neq s_0^m(b - i + 1)\} \\ \leq 2^{-m[H(S_0 | S_i) - bI(X_0, \dots, X_{i-1}; Y_i | X_i, \dots, X_{K-1}) - \epsilon']}, \end{aligned} \quad (23)$$

where  $\epsilon' \triangleq (4 + 6bi)\epsilon$ .

For sources and channels satisfying the conditions of the theorem, by appropriately choosing  $\epsilon$  and letting  $m, n \rightarrow \infty$ , we can have an arbitrarily small probability of error.

## APPENDIX B

### PROOF OF THEOREM 1: SEMI-REGULAR ENCODING AND BACKWARD DECODING

In backward decoding for the single relay channel [6], while the relay decodes each message block right after it is transmitted as in the sliding-window decoding scheme in Section A, the destination waits until all message blocks are transmitted and decodes them in the reverse order by removing the interference from the decoded messages. In backward decoding, each message is decoded at the destination using the signal received in a single block as opposed to signal combining in sliding-window decoding; however, backward decoding introduces additional delay, which grows with the number of relays in the network.

We use the multiple-relay backward decoding scheme for the transmission over the channel [11], [13]. This is a nested backward decoding scheme constructed recursively such that, in each step of the recursion, a new node decodes the messages up to that point using backward decoding.

Rather than the joint decoding approach in Section A, in the case of backward decoding we use separate source and channel encoders/decoders at each node. In the case of separate source-channel coding, the source samples need to be compressed separately for each node such that each node receives enough information to decode the source sequence when combined with its own side information sequence. For example, in the case of a single relay terminal, the source



terminal generates two independent bin indices, one for the relay and one for the destination. The transmission of the bin indices require using channel codes at different rates. However, note that we need a different rate for each receiving node in the network, and all the terminals participating in the transmission of the source terminals to the same node can use the same rate for their codes.

Due to the nested structure of the backward decoding scheme its complexity increases quickly with the increasing number of relays. Hence, for simplicity, we present the transmission scheme for  $K = 2$  relays with a decoding order of  $T_1, T_2, T_3$ . Extension to the setup with more relays is a direct generalization.

As in Section A, for a given  $(n, m)$  pair, we build the code as follows. Fix  $p(x_0, x_1, x_2)$  such that (3) holds. A total of  $B^2 m$  source samples will be transmitted over  $(B + 1)^2 n$  channel uses. This corresponds to a source-channel code rate of  $(B + 1)^2 n / B^2 m$  which gets arbitrarily close to  $n/m$  as  $B \rightarrow \infty$ .

*Source code generation:* Corresponding to each terminal  $T_i$ , for  $i = 1, 2, 3$ , we consider  $M_i = 2^{mR_i}$  bins, called the  $T_i$  bins. All possible source outcomes  $s_0^m \in T_{S_0}^m$  are partitioned randomly and uniformly into these bins, independently for each side information sequence, i.e., the distribution into  $M_i$  bins for  $S_i$  is independent of the distribution into  $M_j$  bins for  $S_j$  for  $i \neq j$ . This bin assignment, which corresponds to source compression, is made available to all the terminals.

*Channel code generation:* For the channel codebook, generate at random  $M_3$  channel codewords  $x_2^n(j_3)$  for  $j_3 \in [1, M_3]$  i.i.d. with  $p(x_2^n(j_3)) = \prod_{t=1}^n p(x_{2,t})$ , and index them as  $x_2^n(j_3)$  with  $j_3 \in [1, M_3]$ .

Then for each  $x_2^n(j_3)$ , generate at random  $M_2$  conditionally independent channel codewords  $x_1^n(j_2|j_3)$ ,  $j_2 \in [1, M_2]$ , with probability  $p(x_1^n(j_2|j_3)) = \prod_{t=1}^n p(x_{1,t}|x_{2,t}(j_3))$ , and index them as  $x_1^n(j_2|j_3)$  with  $j_2 \in [1, M_2]$ .

Finally, generate at random the codebook of size  $M_1$  for each possible combination of  $(x_1^n(j_2|j_3), x_2^n(j_3))$ , with probability  $p(x_0^n(j_1|j_2, j_3)) = \prod_{t=1}^n p(x_{0,t}|x_{1,t}(j_2, j_3), x_{2,t}(j_3))$ , and index them as  $x_0^n(j_1|j_2, j_3)$  with  $j_1 \in [1, M_1]$ .

*Encoding:* Consider a source sequence  $s_0^{B^2 m}$  of length  $B^2 m$ . Partition this sequence into  $B^2$  portions,  $s_{0,b}^m$ ,  $b = 1, \dots, B^2$ . Similarly, partition the side information sequences into  $B^2$  length- $m$  blocks  $s_i^{B^2 m} = [s_{i,1}^m, \dots, s_{i,B^2}^m]$  for  $i = 1, 2, 3$ . The bin index of the  $j$ th block of the source

Block	1	2	...	B	B+1
$T_0$	$x_0^n(w_{1,1} 1, 1)$	$x_0^n(w_{2,1} w_{1,2}, 1)$	...	$x_0^n(w_{B,1} w_{B-1,2}, 1)$	$x_0^n(1 w_{B,2}, 1)$
$T_1$	$x_1^n(1 1)$	$x_1^n(\hat{w}_{1,2}^1 1)$	...	$x_1^n(\hat{w}_{B-1,2}^1 1)$	$x_1^n(\hat{w}_{B,2}^1 1)$
$T_2$	$x_2^n(1)$	$x_2^n(1)$	...	$x_2^n(1)$	$x_2^n(1)$
Block	$(B+1)+1$	$(B+1)+2$	...	$(B+1)+B$	$(B+1)+B+1$
$T_0$	$x_0^n(w_{B+1,1} 1, w_{1,3})$	$x_0^n(w_{B+2,1} w_{B+1,2}, w_{2,3})$	...	$x_0^n(w_{2B,1} w_{2B-1,2}, w_{B,3})$	$x_0^n(1 w_{2B,2}, 1)$
$T_1$	$x_1^n(1 \hat{w}_{1,3}^1)$	$x_1^n(\hat{w}_{B+1,2}^1 \hat{w}_{2,3}^1)$	...	$x_1^n(\hat{w}_{2B-1,2}^1 \hat{w}_{B,3}^1)$	$x_1^n(\hat{w}_{2B,2}^1 1)$
$T_2$	$x_2^n(\hat{w}_{1,3}^2)$	$x_2^n(\hat{w}_{2,3}^2)$	...	$x_2^n(\hat{w}_{B,3}^2)$	$x_2^n(1)$
...	...	...	...	...	...
Block	$k(B+1)+1$	$k(B+1)+2$	...	$k(B+1)+B$	$k(B+1)+B+1$
$T_0$	$x_0^n(w_{kB+1,1} 1, w_{(k-1)B+1,3})$	$x_0^n(w_{kB+2,1} w_{kB+1,2}, w_{(k-1)B+2,3})$	...	$x_0^n(w_{kB+B,1} w_{kB+B-1,2}, w_{(k-1)B+B,3})$	$x_0^n(1 w_{kB+B,2}, 1)$
$T_1$	$x_1^n(1 \hat{w}_{(k-1)B+1,3}^1)$	$x_1^n(\hat{w}_{kB+1,2}^1 \hat{w}_{(k-1)B+2,3}^1)$	...	$x_1^n(\hat{w}_{kB+B-1,2}^1 \hat{w}_{(k-1)B+B,3}^1)$	$x_1^n(\hat{w}_{kB+B,2}^1 1)$
$T_2$	$x_2^n(\hat{w}_{(k-1)B+1,3}^2)$	$x_2^n(\hat{w}_{(k-1)B+2,3}^2)$	...	$x_2^n(\hat{w}_{(k-1)B+B,3}^2)$	$x_2^n(1)$
...	...	...	...	...	...
Block	$B(B+1)+1$	$B(B+1)+2$	...	$B(B+1)+B$	$B(B+1)+B+1$
$T_0$	$x_0^n(1 1, w_{B(B-1)+1,3})$	$x_0^n(1 1, w_{B(B-1)+2,3})$	...	$x_0^n(1 1, w_{B^2,3})$	$x_0^n(1 1, 1)$
$T_1$	$x_1^n(1 \hat{w}_{B(B-1)+1,3}^1)$	$x_1^n(1 \hat{w}_{B(B-1)+2,3}^1)$	...	$x_1^n(1 \hat{w}_{B^2,3}^1)$	$x_1^n(1 1)$
$T_2$	$x_2^n(\hat{w}_{B(B-1)+1,3}^2)$	$x_2^n(\hat{w}_{B(B-1)+2,3}^2)$	...	$x_2^n(\hat{w}_{B^2,3}^2)$	$x_2^n(1)$

Fig. 5. Channel codeword assignment for the semi-regular encoding and backward decoding scheme with  $K = 2$  relays. We transmit  $B^2$  source blocks to the destination in  $(B+1)^2$  channel blocks.

output sequence  $s_{0,j}^m$  with respect to  $T_i$  bins is denoted by  $w_{j,i}$ . The estimate of  $w_{j,i}$  at node  $k$ ,  $k = 1, \dots, K+1$ , is denoted by  $\hat{w}_{j,i}^k$ . See Fig. 5 for an illustration of the encoding scheme.

In block 1,  $T_0$  observes  $s_{0,1}^m$ , and finds the corresponding bin index  $w_{1,1} \in [1, M_1]$ . It transmits the channel codeword  $x_0^n(w_{1,1}|1, 1)$ . The relays  $T_1$  and  $T_2$  simply transmit  $x_1^n(1|1)$  and  $x_2^n(1)$ , respectively. In block 2,  $T_0$  transmits the channel codeword  $x_0^n(w_{2,1}|w_{1,2}, 1)$ . The relays  $T_1$  and  $T_2$  transmit  $x_1^n(\hat{w}_{1,2}^1|1)$  and  $x_2^n(1)$ , respectively, where  $\hat{w}_{1,2}^1$  is the  $T_2$  bin index of the estimate  $\hat{s}_{1,1}^m$  at the relay  $T_1$ . In the following blocks  $b = 2, \dots, B$ , the source terminal transmits the channel codeword  $x_0^n(w_{b,1}|w_{b-1,2}, 1)$  where  $w_{b,i} \in [1, M_i]$  for  $i = 1, 2$ . In block  $B+1$ ,  $T_0$  transmits  $x_0^n(1|w_{2,B}, 1)$ .

The first relay  $T_1$  estimates the source block  $s_{0,b-1}^m$  at the end of block  $b-1$ , denoted by  $\hat{s}_{1,b-1}^m$ , and finds the corresponding  $T_2$  bin index  $\hat{w}_{b-1,2}^1 \in [1, M_2]$ . At block  $b$ , for  $b = 2, \dots, B+1$ ,  $T_1$  transmits the channel codeword  $x_1^n(\hat{w}_{b-1,2}^1|1)$ . For the first  $B+1$  channel blocks terminal  $T_2$  transmits  $x_2^n(1)$ .

At the end of channel block  $B+1$ , the relay  $T_2$  decodes the first source block  $s_{0,1}^m$  by backward decoding. Having estimated the source blocks  $s_{0,1}^m, \dots, s_{0,B}^m$  by backward decoding,  $T_2$  joins the transmission for forwarding the  $T_3$  bin indices of the first source block to the destination.

In the following channel blocks  $T_0$  superposes the transmission of the second source block

to  $T_1$  and  $T_2$  on the transmission of the  $T_3$  bin indices of the first source block to  $T_3$ . In the channel blocks from  $B + 1$  up to  $2B + 1$ , the source terminal transmits, in order,

$$x_0^n(w_{B+1,1}|1, w_{1,3}), \dots, x_0^n(w_{2B,1}|w_{2B-1,2}, w_{B,3}), x_0^n(1|w_{2B,2}, 1).$$

The relay  $T_1$  transmits, in order,

$$x_1^n(1|\hat{w}_{1,3}^1), x_1^n(\hat{w}_{B+1,2}^1|\hat{w}_{2,3}^1), \dots, x_1^n(\hat{w}_{2B-1,2}^1|\hat{w}_{B,3}^1), x_1^n(\hat{w}_{2B,2}^1|1).$$

Finally, the second relay  $T_2$  transmits  $x_2^n(\hat{w}_{1,3}^2), \dots, x_2^n(\hat{w}_{B,3}^2, x_2^n(1))$ .

They continue similarly for a total of  $B$  channel blocks of  $Bn$  channel uses each. In each of these  $B$  groups of channel blocks  $Bm$  source samples are encoded by terminal  $T_0$ . In the last block of  $(B + 1)n$  channel uses, no new source samples are encoded. Terminal  $T_0$  transmits

$$x_0^n(1|1, w_{B(B-1)+1,3}), \dots, x_0^n(1|1, w_{B^2,3}), x_0^n(1|1, 1),$$

while  $T_1$  transmits  $x_1^n(1|\hat{w}_{B(B-1)+1,3}^1), \dots, x_1^n(1|\hat{w}_{B^2,3}^1), x_1^n(1|1)$ , and having estimated the  $T_3$  bin indices for the last  $Bm$  source samples,  $T_2$  transmits  $x_2^n(\hat{w}_{B(B-1)+1,3}^2), \dots, x_2^n(\hat{w}_{B^2,3}^2), x_2^n(1)$ . It can be noted that the last channel block of  $n$  channel uses is unused, but is included to simplify the expressions as it does not reduce the source-channel code rate in the limit of an infinite number of source and channel blocks.

*Decoding and error probability analysis:* Since each node tries to decode each block of the source sample, we denote the estimate of source block  $s_{0,b}^m$  at node  $T_k$ ,  $k = 1, \dots, K + 1$ , by  $\hat{s}_{1,b}^m$ .

The relay  $T_1$  decodes the source signal by sequentially reconstructing source block  $s_{0,b}^m$  at the end of the channel block, in which the corresponding  $T_1$  bin index is transmitted by the source terminal  $T_0$ .

Consider channel block  $k(B + 1) + b$  for  $k = 0, 1, \dots, B - 1$  and  $b = 1, \dots, B$ . Assume that at the end of block  $k(B + 1) + b - 1$ ,  $T_1$  has estimated the source blocks  $s_{0,1}^m, \dots, s_{0,kB+b-1}^m$  correctly, i.e.,  $\hat{s}_{1,b}^m = s_{0,b}^m$  for  $b = 1, \dots, kB + b - 1$ . Hence, it can correctly find the  $T_2$  bin index for  $s_{0,kB+b-1}^m$  and the  $T_3$  bin index for  $s_{0,(k-1)B+b}^m$ , i.e.,  $\hat{w}_{kB+b-1,2}^1 = w_{kB+b-1,2}$  and  $\hat{w}_{(k-1)B+b,3}^1 = w_{(k-1)B+b,3}$ . Using this information and its received signal  $y_1^n(k(B + 1) + b)$ , the  $T_1$  channel decoder attempts to decode  $w_{kB+b,1}$ , i.e., the  $T_1$  bin index corresponding to  $s_{0,kB+b}^m$ , by looking for a unique index  $w$  such that

$$(x_0^n(w|w_{kB+b-1,2}, w_{(k-1)B+b,3}), x_1^n(\hat{w}_{kB+b-1,2}^1|\hat{w}_{(k-1)B+b,3}^1), x_2^n(\hat{w}_{(k-1)B+b,3}^2), Y_1^n(b)) \in A_\epsilon^n(X_1, X_2, X_3, Y_1),$$

where  $\hat{w}_{b,3}^i = w_{b,3} = 1$  for  $i = 1, 2$ , if  $b < 1$ . If such a unique index is found, then it is assigned as the bin index estimation of the source sequence  $s_{0,b}^m$  at node  $T_1$ ,  $w_{b,1}^1$ .

The estimated bin index is then given to the  $T_1$  source decoder. With the  $T_1$  bin index and the side information  $s_{1,kB+b}^m$ , the  $T_1$  source decoder estimates  $s_{0,kB+b}^m$  by looking for a unique source codeword which is jointly typical with  $S_1^m(kB+b)$  and whose bin index is  $\hat{w}_{kB+b,1}^1$ . If such a unique codeword is found, it is assigned as the source estimation at  $T_1$ , denoted by  $\hat{s}_{1,b}^m$ .

At the end of channel block  $k(B+1) + B + 1$ , for  $k = 0, \dots, B-1$ , the relay  $T_2$  decodes the  $T_2$  bin index for the source block  $s_{0,kB+B}^m$ . It then continues decoding in the reverse order by backward decoding. Assuming that it has estimated the source blocks  $s_{0,1}^m, \dots, s_{0,(k-1)B+B}^m$  and  $s_{0,kB+b+1}^m, \dots, s_{0,kB+B}^m$  correctly, it can find the  $T_1$  bin index of  $s_{0,kB+b+1}^m$  and the  $T_3$  bin index of  $s_{0,(k-1)B+b}^m$ . Using this information and its received signal  $Y_2^n(k(B+1) + b)$ , the  $T_2$  channel decoder attempts to decode  $w_{kB+b,2}$  by looking for a unique index  $w$  such that

$$\begin{aligned} (x_0^n(w_{kB+b+1,1}|w, w_{(k-1)B+b,3}), x_1^n(w|\hat{w}_{(k-1)B+b,3}^1), x_2^n(\hat{w}_{(k-1)B+b,3}^2), Y_2^n(k(B+1) + b)) \\ \in A_\epsilon^n(X_1, X_2, X_3, Y_2). \end{aligned}$$

If such a unique index is found, then it is assigned as the bin index estimation of the source sequence  $s_0^m(kB+b)$  at node  $T_2$ ,  $\hat{w}_{kB+b,2}$ .

The estimated bin index is then given to the  $T_2$  source decoder. With the  $T_2$  bin index and the side information  $s_{2,kB+b}^m$ , the  $T_2$  source decoder estimates  $s_{0,kB+b}^m$  by looking for a unique source codeword which is jointly typical with  $S_2^m(kB+b)$  and whose bin index is  $\hat{w}_{kB+b,2}^2$ . If such a unique codeword is found, it is assigned as the source estimation at  $T_1$ , denoted by  $\hat{s}_{2,kB+b}^m$ .

Decoding at the destination node  $T_3$  is also done using backward decoding, but the destination waits till the end of channel block  $B(B+1) + B$ . It first tries to decode the  $T_3$  bin index of the last source block  $s_{0,B^2}^m$  using the received signal at channel block  $B(B+1) + B$ . Consider decoding of  $s_0^m((k-1)B+b)$  at channel block  $k(B+1) + b$  for  $k = 1, \dots, B$  and  $b = 1, \dots, B$ . Assuming that it has decoded the source blocks  $s_{0,(k-1)B+b+1}^m, \dots, s_{0,B^2}^m$  correctly, it can find the  $T_1$  bin index of  $s_{0,kB+b}^m$  and the  $T_2$  bin index of  $s_{0,kB+b-1}^m$ . Using these information and its received signal  $Y_3^n(k(B+1) + b)$ , the  $T_3$  channel decoder attempts to decode  $w_{(k-1)B+b,3}$  by looking for a unique index  $w$  such that

$$(x_0^n(w_{kB+b,1}|w_{kB+b-1,2}, w), x_1^n(\hat{w}_{kB+b-1,2}^1|w), x_2^n(w), Y_3^n(k(B+1) + b))$$

$$\in A_\epsilon^n(X_1, X_2, X_3, Y_3).$$

If such a unique index is found, then it is assigned as the bin index estimation of the source sequence  $s_0^m((k-1)B+b)$  at node  $T_2$ ,  $\hat{w}_{kB+b,2}$ .

The estimated bin index is then given to the  $T_3$  source decoder. With the  $T_3$  bin index and the side information  $s_{3,(k-1)B+b}^m$ , the  $T_3$  source decoder estimates  $s_{0,(k-1)B+b}^m$  by looking for a unique source codeword that is jointly typical with  $S_3^m((k-1)B+b)$  and whose bin index is  $\hat{w}_{(k-1)B+b,3}^3$ .

We upper bound the probability of error by counting errors made not only at the destination, but at any terminal in the network.

$$P_e \leq \sum_{s^{B^2m}} \Pr \left\{ \bigcup_{b=1,\dots,B^2} \bigcup_{k=1,\dots,K+1} \{ \hat{S}_{b,k}^m \neq s_{0,b}^m | S^{B^2m} = s^{B^2m} \} \right\} Pr \{ S^{B^2m} = s^{B^2m} \}. \quad (24)$$

We can write the above union in a recursive manner by considering the decoding order of the backward decoding scheme; that is, the above event of having an error in estimating any of the source blocks at any of the nodes can be written as the union of error events, such that each corresponds to an error event at one stage of decoding given that no error has been made previously. Note that the decoding order is as follows:  $\hat{s}_{1,1}^m, \hat{s}_{1,2}^m, \dots, \hat{s}_{1,B}^m, \hat{s}_{2,B}^m, \dots, \hat{s}_{2,1}^m, \hat{s}_{1,B+1}^m, \dots, \hat{s}_{1,2B}^m, \hat{s}_{2,2B}^m, \dots, \hat{s}_{2,B+1}^m, \dots, \hat{s}_{1,(B-1)B+1}^m, \dots, \hat{s}_{1,B^2}^m, \hat{s}_{2,B^2}^m, \dots, \hat{s}_{2,(B-1)B+1}^m, \hat{s}_{3,B^2}^m, \hat{s}_{3,B^2-1}^m, \dots, \hat{s}_{3,1}^m$ .

We have

$$P_e \leq \sum_{s^{B^2m}} \left\{ \sum_{b=1,\dots,B^2} \sum_{k=1,\dots,K+1} \Pr \{ \hat{S}_{b,k}^m \neq s_{0,b}^m | E_{b,k}, S^{B^2m} = s^{B^2m} \} \cdot \Pr \{ E_{b,k} | S^{B^2m} = s^{B^2m} \} \right\} Pr \{ S^{B^2m} = s^{B^2m} \}, \quad (25)$$

where we define  $E_{b,k}$  as the event that all previous estimations are correct when  $T_k$  estimates  $S_{0,b}^m$ . Then we get

$$P_e \leq \sum_{s^{B^2m}} \left\{ \sum_b \sum_k \Pr \{ \hat{S}_{b,k}^m \neq s_{0,b}^m | E_{b,k}, S^{B^2m} = s^{B^2m} \} \right\} \cdot Pr \{ S^{B^2m} = s^{B^2m} \}, \quad (26)$$

$$\leq \sum_{s^{B^2m}} \left\{ \sum_b \sum_k \Pr \{ \hat{S}_{b,k}^m \neq s_{0,b}^m, \hat{w}_{b,k}^k = w(b,k) | E_{b,k}, S^{B^2m} = s^{B^2m} \} \right. \\ \left. + \Pr \{ \hat{S}_{b,k}^m \neq s_{0,b}^m, \hat{w}_{b,k}^k \neq w(b,k) | E_{b,k}, S^{B^2m} = s^{B^2m} \} \right\} \cdot Pr \{ S^{B^2m} = s^{B^2m} \}, \quad (27)$$

$$\leq \sum_{s^{B^2m}} \left\{ \sum_b \sum_k \Pr\{\hat{S}_{b,k}^m \neq s_{0,b}^m | \hat{w}_{b,k}^k = w(b,k), E_{b,k}, S^{B^2m} = s^{B^2m}\} \right. \\ \left. + \Pr\{\hat{w}_{b,k}^k \neq w(b,k) | E_{b,k}, S^{B^2m} = s^{B^2m}\} \right\} \cdot \Pr\{S^{B^2m} = s^{B^2m}\}, \quad (28)$$

$$= \sum_b \sum_k \Pr\{\hat{S}_{b,k}^m \neq s_{0,b}^m | \hat{w}_{b,k}^k = w(b,k), E_{b,k}\} \\ + \sum_{s^{B^2m}} \left\{ \sum_b \sum_k \Pr\{\hat{w}_{b,k}^k \neq w(b,k) | E_{b,k}, S^{B^2m} = s^{B^2m}\} \right\} \Pr\{S^{B^2m} = s^{B^2m}\}, \quad (29)$$

$$= \sum_b \sum_k \Pr\{\hat{S}_{b,k}^m \neq s_{0,b}^m | \hat{w}_{b,k}^k = w(b,k), E_{b,k}\} \\ + \sum_b \sum_k \sum_{w(b,k) \in \{1, \dots, 2^{mR_k}\}} \Pr\{\hat{w}_{b,k}^k \neq w(b,k) | E_{b,k}, W_{b,k} = w(b,k)\} \Pr\{W_{b,k} = w(b,k)\} \\ = \sum_b \sum_k \Pr\{\hat{S}_{b,k}^m \neq s_{0,b}^m | \hat{w}_{b,k}^k = w(b,k), E_{b,k}\} \\ + \sum_b \sum_k \frac{1}{2^{mR_k}} \sum_{w(b,k) \in \{1, \dots, 2^{mR_k}\}} \Pr\{\hat{w}_{b,k}^k \neq w(b,k) | E_{b,k}, W_{b,k} = w(b,k)\}. \quad (31)$$

Now, note in (31) that each term in the first summation corresponds to the error event at the source decoder of  $T_k$  given that it is provided with the correct bin index, and each term in the second summation corresponds to the error event at the channel decoder of  $T_k$ , both conditioned on the fact that all the estimations up to that instant are correct. Following the usual arguments, we get

$$P_{e,i}(b) \leq \sum_b \sum_k \left\{ 2^{mR_k} \cdot 2^{m(H(S_0|S_k)+\epsilon)} + 2^{mR_k} \cdot 2^{-n(I(X_0^{k-1}; Y_k | X_k^{K+1})+\epsilon)} \right\} \quad (32)$$

Setting  $R_k = H(S_0|S_k)$ , and letting  $m, n \rightarrow \infty$  while  $\epsilon \rightarrow 0$ , the error probability vanishes under the assumptions of the theorem.

## REFERENCES

- [1] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. on Information Theory*, vol. 25, no. 5, pp. 572 – 584, September 1979.
- [2] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian-Wolf: Theory, algorithms and scaling laws," *IEEE Trans. Information Theory*, vol. 51, no. 12, pp. 4057–4073, Dec. 2005.
- [3] H. Behroozi and M. Soleymani, "Sending correlated Gaussian sources over a Gaussian MAC: To code, or not to code," in *Proc. IEEE Int'l Conf. on Communication (ICC)*, Glasgow, Scotland, Jun. 2007.
- [4] I. E. Aguerri and D. Gündüz, "Hybrid digital-analog transmission for the gaussian one-helper problem," in *Proc. IEEE Global Communications Conf. (GLOBECOM)*, Miami, FL, Dec. 2010.

- [5] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Information Theory*, vol. 28, no. 6, p. 841850, Nov. 1982.
- [6] F. M. J. Willems, *Information Theoretical Results for the Discrete Memoryless Multiple Access Channel*. Leuven, Belgium: PhD dissertation, Katholieke Univ. Leuven, Oct. 1982.
- [7] J. N. Laneman and G. Kramer, "Window decoding for the multiple access channel with generalized feedback," in *Proc. IEEE Int'l Symposium on Information Theory*, Chicago, IL, Jun. 2004.
- [8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [9] M. R. Aref, *Information Flow in Relay Networks*. Stanford, CA: Ph.D. dissertation, Stanford Univ., 1980.
- [10] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Information Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.
- [11] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. Fuja, "Bandwidth- and power-efficient routing in linear wireless networks," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2624–2633, Jun. 2006.
- [12] P. Razaghi and W. Yu, "Parity forwarding for multiple-relay networks," *IEEE Trans. Information Theory*, vol. 55, no. 1, pp. 158–173, Jan. 2009.
- [13] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. 41st Annual Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Oct. 2003.
- [14] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, Sep. 1973.
- [15] D. Gündüz and E. Erkip, "Reliable cooperative source transmission with side information," in *Proc. IEEE Information Theory Workshop (ITW)*, Bergen, Norway, Jul. 2007.
- [16] R. Kwak, W. Lee, A. E. Gamal, and J. Cioffi, "Relay with side information," in *IEEE Int'l Symposium on Information Theory (ISIT)*, Nice, France, Jun. 2007.
- [17] B. Smith and S. Vishwanath, "Cooperative communication in sensor networks: Relay channels with correlated sources," in *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2004.
- [18] M. Sefidgaran, B. Akhbari, Y. Mohsenzadeh, and M. R. Aref, "Source transmission over relay channel with correlated relay side information," in *Proc. IEEE Int'l Symposium on Information Theory*, Seoul, South Korea, Jul. 2009.
- [19] D. Gündüz, E. Erkip, A. Goldsmith, and H. V. Poor, "Cooperative relaying in sensor networks," in *Proc. Int'l Conference on Cognitive Radio Oriented Wireless Networks and Communications*, Cannes, France, Jun. 2010.
- [20] —, "Lossy source transmission over the relay channel," in *Proc. IEEE Int'l Symposium on Information Theory*, Toronto, ON, Canada, Jul. 2008.
- [21] —, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. Information Theory*, vol. 55, no. 9, pp. 3927–3944, Sep. 2009.
- [22] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [23] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.
- [24] S. Shamai and S. Verdú, "Capacity of channels with uncoded side information," *European Transactions on Telecommunications and Related Technologies*, vol. 6, no. 5, pp. 587–600, Sept.-Oct. 1995.

- [25] E. Tuncel, “Slepian-Wolf coding over broadcast channels,” *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.
- [26] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York, NY: Academic Press, 1981.
- [27] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos, “Source-channel diversity for parallel channels,” *IEEE Trans. Information Theory*, vol. 51, no. 10, pp. 3518–3539, Oct. 2005.