Compress-and-Forward Relaying Based on Symbol-Wise Joint Source-Channel Coding

©2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

RICARDO BLASCO-SERRANO, RAGNAR THOBABEN, AND MIKAEL SKOGLUND

Stockholm 2010

Communication Theory Department
School of Electrical Engineering
KTH Royal Institute of Technology

IR-EE-KT 2010:011
Compress-and-Forward Relaying Based on Symbol-Wise Joint Source-Channel Coding

Ricardo Blasco-Serrano, Ragnar Thobaben, and Mikael Skoglund
School of Electrical Engineering
Royal Institute of Technology (KTH) and ACCESS Linnaeus Centre
SE-100 44 Stockholm, Sweden
Email: {ricardo.blasco, ragnar.thobaben, mikael.skoglund}@ee.kth.se

Abstract—We propose a new compress-and-forward implementation for the relay channel based on joint source-channel coding techniques. The relay performs scalar quantization of its observation in combination with a redundant index mapping. Our system utilizes the correlation between the quantized signal and the direct-link observation of the transmitted symbols as redundancy for error protection on the relay-to-destination link. In order to fully exploit this correlation the destination requires iterative decoding to recover the quantized observation sent by the relay. Once regenerated, this quantized signal is optimally combined with the direct-link observation to decode the message conveyed by the source. By quantizing the observed signal itself rather than a measure on the reliability of the information bits (e.g. a posteriori probabilities from a decoder), and by using digital communication methods on the relay-to-destination link our system yields superior performance to that of amplify-and-forward, decode-and-forward and previous implementations of compress-and-forward based on soft decoding.

I. INTRODUCTION

Cooperation is one of the present challenges in digital communications. Potential benefits are for example higher diversity or rate, helping to extend range or capacity of a communication network respectively. Several families of protocols for the classical cooperative model known as the relay channel exist. Two of the most significant ones are decode-and-forward (DF) and compress-and-forward (CF) [1]. Roughly speaking, in DF a relay node tries to decode the message from the source, reencode (possibly a compressed version of) it and forward this information to the destination. DF is usually implemented using distributed channel coding schemes (e.g. turbo codes [2] or LDPC codes [3]). Clearly, the assumption of decodability at the relay restricts the scenarios where DF is useful. In contrast, in CF the relay listens to the message from the source and sends a compressed version of its observation to the receiver. Most of the work on practical implementations has focused on DF whereas only a few schemes for CF have been presented so far [4], [5] and [6].

In this paper we propose a new implementation for CF relaying. Our approach is based on joint source-channel coding (JSCC) ideas. In our system the relay listens to the source transmission and performs scalar quantization on the received data. We argue that in the scenarios where DF is not possible

the relay node should provide the destination with information about its observation of the source signal (i.e. transmitted symbols) rather than about the message (i.e. information bits). Instead of removing the correlation between the quantized version of the relay observation and the signal received at the destination through the direct link [7], we propose to exploit it as redundancy for channel-error protection. Iterative decoding at the receiver becomes necessary in order to regenerate the JSCCed quantized observation of the relay. The source message is finally recovered using this reconstruction and the direct observation of the transmitted message, benefiting from the relay cooperation.

This paper is organized as follows: in Section II we introduce the system model and discuss previously presented implementations of CF. In Section III we present our new scheme for CF relaying and the design guidelines. Performance results and comparison to other alternatives are presented in Section IV. Section V concludes our work.

II. BACKGROUND

A. Transmission model

In the scenario considered here, depicted in Figure 1, the source tries to convey a sequence of equally likely i.i.d. bits $\mathbf{U}$ to the destination. Communication is divided into two phases. In the first one the source performs channel coding ($\mathbf{U}_s$ with rate $R_s$) and modulation on the information bits, and transmits the resulting symbols $\mathbf{X}_S$. Both relay and destination observe noisy versions of this transmitted signal, $\mathbf{Y}_{SR}$ and $\mathbf{Y}_{SD}$ respectively. In the second phase the relay processes its observation prior to transmission to the destination. In doing so a relay based on CF introduces some distortion with the aim of discretizing the observation (or some signal derived from it). In general, this quantized observation can be compressed further taking into account that the destination already has some side information about it available from the source-to-destination (SD) link reception $\mathbf{Y}_{SD}$ [7].

Alternatively, the relay could simply retransmit its observation with some scaling factor to account for power constraints. This simple strategy is known as amplify-and-forward (AF) and can be seen as the simplest realization of CF.

Finally, using the noisy observation of the relay transmission $\mathbf{Y}_{RD}$ as well as $\mathbf{Y}_{SD}$, the destination tries to recover the information bits $\mathbf{U}$.
Throughout this paper transmissions over the different channels are assumed to be interference-free (i.e. the channels are orthogonal). Vectors are represented with bold face $x$, scalars with normal face $y$, and the element in position $i$ of a vector $x$ is denoted as $x(i)$ or $x_i$. In what follows log denotes the natural logarithm.

B. Related work

Several different implementations of CF have been proposed in the literature. They can roughly be divided into two groups depending on the strategy adopted at the relay.

The first group is represented by Figure 2(a). Relays in this class apply some loss compression technique (e.g., quantization) directly to their observation $Y_{SR}$. The resulting signal $Y_Q$ is further processed in order to perform lossless compression, protect it from errors in the channel, etc. The receiver, using both its observations $Y_{SD}$ and $Y_{RD}$, tries to regenerate $Y_Q$ and combines this reconstruction optimally with $Y_{SD}$ prior to decoding. The ultimate performance of this method relies essentially on two things. The ability of the relay to introduce as little distortion as possible in the lossy compression stage and the use of appropriate techniques to exploit the side information contained in $Y_{SD}$ (i.e., Wyner-Ziv coding [7]). Moreover, both aspects are clearly coupled: small distortions are only possible if redundancy is fully exploited and the signal-to-noise ratio (SNR) on the RD link is high enough. For poor RD links reducing the distortion will require from bandwidth expansion. An example in this class for non-orthogonal transmissions from the source and the relay is presented in [4].

On the other hand, relay implementations belonging to the second group (see Figure 2(b)) first perform channel decoding using some soft-input/soft-output (SISO) algorithm to obtain log-likelihood ratios (LLRs) on the information bits $L_{U(k)} = \log(p(U(k) = 1|Y_{SR})/p(U(k) = 0|Y_{SR}))$. These LLRs are real numbers and need to be discretized using lossy compression methods if a digital modulation is to be used over the RD link. Examples falling in this group are [5] and [6]. The main advantage of this relaying scheme is that by decoding, compression is implicitly performed. For example, if the source is using a channel code with rate $R_x = 1/M$ for every $M$ channel observations only one LLR is generated. If both communication phases consist of the same number of channel uses each LLR can be discretized using $M$ bits. For low code rates unnoticeable distortions can be obtained without incurring in bandwidth expansion.

However, with increasing capacity on the RD link only algorithms in the first group can approach the performance of a distributed antenna system (DAS) in which the receiver has access to both $Y_{SD}$ and $Y_{RD}$. The reason for this is that, even under the assumption of high capacity RD link, algorithms based on processing the LLRs would only provide these LLRs (with some negligible distortion) to the receiver. Decoders from [5] and [6] use the LLRs as side information resulting in a performance loss with respect to the optimal performance. Relaying schemes based on compression of $Y_{SR}$ directly could convey the observation under the high-capacity RD link assumption. This would allow the decoder to combine both observations optimally and achieve the same performance of a DAS.

C. Bandwidth expansion

From the discussion above it is clear that the performance of CF algorithms is highly dependent on the rate-distortion tradeoff in the relay design. The lower the distortion the larger the bit rate required on the RD link and vice versa. Moreover, whenever the RD link has a low SNR, reduction of distortion leads inexorably to a larger bandwidth requirement. The potential of CF with bandwidth expansion (BE) in cellular networks has been recently pointed out in [8]. Our work focuses on scenarios where the SNR on the RD link is low to moderate so that digital transmission outperforms analog methods (i.e., CF as opposed to AF). Therefore we make use of BE and low order modulations at the relay (i.e., BPSK). The system to be presented here works of course with higher SNR and larger constellations that avoid the necessity for BE. However, under these conditions AF shows reasonably good performance, making it hard to motivate the use of computationally more complex techniques such as iterative decoding in order to obtain only marginal gains.

III. SYSTEM DESIGN

We propose a new CF relaying scheme based on lossy compression of the relay observation. Its operation is based on...
symbol-wise joint source-channel coding and requires iterative decoding. As stated before, the signal resulting from quantizing the observation $Y_{SR}$ is correlated with $Y_{SD}$. Conventional CF schemes make use of distributed source coding tools to eliminate this superfluous information and then perform channel coding to combat errors on the RD link by introducing controlled redundancy. Instead of this, we propose to keep the correlation and process it properly so that the decoder can use it in combination with $Y_{SD}$ to combat the effect of noise on the RD link. Relay and destination nodes are implemented as described in the following.

A. Joint source-channel coding at the relay

The relay node, depicted in Figure 3, consists of two stages here named as source code and channel code separated by an interleave (denoted as $\pi$ in the figures). The first one implements a scalar quantizer (SQ) with $m$ bits which uses a redundant index mapping to obtain good distance properties, allowing for iterative decoding [9]. To realize this we concatenate the SQ with an outer block code, $C_{i}$, with rate $R_{o} = \frac{m}{2^t} < 1$. The second stage presents an inner code $C_{i}$ with rate $R_{o}$ concatenated with a puncturer. This is used to remove unnecessary redundancy, adapting the rate to the channel conditions on the RD link. Alternatively, one could choose to reduce the amount of data to be transmitted from $m$ bits/sample to $\frac{1}{2}H(Y_{Q} \mid Y_{SD})$. The resulting data rate would be controlled by choosing a puncturing rate $R_{p}$ which would have to satisfy $\frac{m}{2^t}R_{o} + \frac{1}{2}H(Y_{Q} \mid Y_{SD})$ for given $R_{i}$, $R_{o}$ and $m$ with the channel capacity $C_{RD}$ on the RD link. However, in this case it would be necessary to add new redundancy to mitigate the effects of the errors on the RD link.

![Fig. 3. Relay structure.](image)

B. Information combining at the destination

The destination (see Figure 4), using both its observations $Y_{SD}$ and $Y_{RD}$, tries to regenerate $Y_{Q}$ with iterative decoding as described below. Once the estimate $\hat{Y}_{Q}$ has been formed, it is optimally combined with $Y_{SD}$ and then used to decode the information bits $\hat{U}$. Given the two independent observations $Y_{Q}$ and $Y_{SD}$ their optimal combination $L_{Y}^{CF}$ is simply the addition of the corresponding LLRs,

$$L_{Y}^{CF} = L_{Y}^{Q} + L_{Y}^{SD}$$

with

$$L_{Y}^{Q} = \log \left( \frac{p(Y_{Q} \mid b = 1)}{p(Y_{Q} \mid b = 0)} \right)$$

and $L_{Y}^{SD}$ defined similarly. Equations (1) and (2) are justified by the fact that $Y_{SR}$ and $Y_{SD}$ are conditionally independent given $X_{S}$; i.e. we have

$$p(Y_{Q}, Y_{SD} \mid X_{S}) = p(Y_{SD} \mid X_{S})p(Y_{Q} \mid X_{S}),$$

or by assuming memoryless channels

$$p(Y_{Q}, Y_{SD} \mid X_{S}) = p(Y_{SD} \mid X_{S})p(Y_{Q} \mid X_{S}).$$

![Fig. 4. Receiver structure at the destination.](image)

C. Iterative decoding

To describe the decoding process at the destination it is necessary to take into account that the output of the SQ has three equivalent representations. It can be identified as a sequence of reconstruction points $Y_{Q} = [Y_{Q}(0) \ Y_{Q}(1) \ldots]$ or as a sequence of cell indices (i.e. a vector of integers) $J = [j_{0} \ j_{1} \ldots]$. Alternatively, it can be represented by a sequence of bits $b$ with each cell number represented by $M$ bits using some unambiguous mapping (e.g. Gray, binary, etc.). That is, $Y_{Q}(n) = j_{n} = j_{n+1} = b_{n,0} b_{n,1} \ldots b_{n,M-1}$. In the latter case the representation takes also into account the redundancy added by $C_{i}$.

Decoding at the destination is performed in an iterative fashion as shown in Figure 4. Initially the SISO decoder for $C_{i}$ obtains extrinsic information $L_{e}^{i}$ about the bits $b$ with no a priori information coming from the other decoder [10]. These extrinsic LLRs are de-interleaved and used by the SISO decoder for $C_{o}$ as a priori information. Using this information and the observation $Y_{SD}$ the decoder generates extrinsic information $L_{e}^{c}$ on the bits $b$ that will allow the SISO decoder for $C_{o}$ to start a new iteration. Success in decoding relies on extracting appropriately the information about $b$ contained in $Y_{SD}$. For this purpose the optimal processing is to calculate the a posteriori LLRs as

$$L_{b_{n,k}}^{opt} = \log \left( \frac{\sum_{j_{n} = b_{n,k}} p(j_{n} \mid Y_{SD}, L_{e}^{c})}{\sum_{j_{n} \neq b_{n,k}} p(j_{n} \mid Y_{SD}, L_{e}^{c})} \right)$$

where “$j \rightarrow b_{n,k} = 0$” and “$j \rightarrow b_{n,k} = 1$” denote the indices $j$ such that the mapping used has bit $b_{n,k}$ equal to 0 or 1 respectively. However, the presence of the code $C_{e}$ at the source, and the interleaver and the channel code at the
relay generates a statistical dependency between \( j_n, Y_{SD} \) and \( L^i_c \) that makes computation of (6) unfeasible due to its high complexity.

We introduce a suboptimal algorithm that works on a symbol-by-symbol basis disregarding partially the statistical dependency between \( j_n, Y_{SD} \) and \( L^i_c \). This assumption is reinforced by the channel conditions for CF (low SNRs) and the presence of a SQ at the relay node which operates on a sample-by-sample basis. In each iteration our suboptimal decoder for \( e_n \) generates the extrinsic information as [10]

\[
L^e_c = L^b_c - L^i_c
\]  

\( (7) \)

Assuming \( j_n \equiv b_n \) resulting from the application of the SQ and \( e_n \) to the observation \( Y_{SR}(n) \) we have

\[
L_{b_n,k} = \log \left( \frac{\sum_{j_n=b_n,k=1}^{1} p(j_n|Y_{SD}(n), L^i_c(n))}{\sum_{j_n=b_n,k=0}^{1} p(j_n|Y_{SD}(n), L^i_c(n))} \right)
\]  

\( (8) \)

where \( L^i_c(n) \) is a vector containing the extrinsic values \( [L^i_c(0), \ldots, L^i_c(M-1)] \) associated to the \( M \) bits \( b_n \) corresponding to the quantization of \( Y_{SR}(n) \), \( Y_Q(n) \). The elements in the summations can be expressed as

\[
p(j_n|Y_{SD}(n), L^i_c(n)) = \frac{1}{c} p(j_n)p(Y_{SD}(n)|j_n)p(L^i_c(n)|j_n)
\]  

\( (9) \)

with

\[
p(L^i_c(n)|j_n) = \prod_{k=0}^{M-1} p(L^i_{c,n,k}|b_{n,k}) = \prod_{k=0}^{M-1} \frac{b_{n,k} L^i_{c,n,k}}{1 + e L^i_{c,n,k}}
\]  

\( (10) \)

where \( c \) is a normalization constant that cancels out in the expression of the LLR. Equation (9) results by assuming conditional independence of \( Y_{SD}(n) \) and \( L^i_c(n) \) given \( j_n \). Information about index \( j_n \) present in \( Y_{SD}(n) \) corresponds to the second factor in (9) and can be expressed as

\[
p(Y_{SD}(n)|j_n) = \frac{1}{c'} \sum_{\{x\}} p(Y_{SD}(n), j_n, x)
\]  

\( (11) \)

\[
= \frac{1}{c'} \sum_{\{x\}} p(Y_{SD}(n)|x)p(j_n|x)p(x)
\]  

\( (12) \)

\[
= \frac{1}{c''} \sum_{\{x\}} p(Y_{SD}(n)|x)p(j_n|x).
\]  

\( (13) \)

The summations run over all possible transmitted values (e.g. ±1 for BPSK). Again \( c' \) and \( c'' \) are normalization constants that have no effect in (8). Equation (12) comes from the conditional independence of the observations given the transmitted symbol and that \( j_n \) is determined by \( Y_{SR}(n) \) and the quantizer design. Equation (13) is obtained under the assumption of equally likely transmitted symbols from the source.

As described before both decoders exchange extrinsic information on the bits rather than on the indices. It is worth remarking that different bit mappings yield different behaviors and performances of the SISO \( e_n \) component block during iterative decoding. The effects of different mappings can be predicted using their EXIT functions [11]. The mapping to be used remains as a degree of freedom for the designer and should be chosen following design criteria for iterative decoding, e.g. [12].

### IV. Simulation results

We present here the results from a series of simulations in order to evaluate the performance of our proposed scheme. All systems shown implement resolution-constrained SQ designed to minimize the distortion of the reconstruction using the Lloyd algorithm [13]. The parameters of the JSCC scheme are obtained following the design principles based on EXIT charts [12]. Binary mapping for SQ outputs is used in all simulations, and the redundant index mapping is generated using a single parity-check code as \( e_n \). \( e_n \) corresponds to a recursive convolutional code with rate \( R_i = 1 \) and generator polynomial \((3/7)\). The interleaver between source and channel codes at the relay is chosen randomly. The additional bits added by the relay are removed using pseudo-random puncturing with the random seed known to both relay and destination. Source messages consist of \( 10^4 \) uniform i.i.d. bits protected with a recursive systematic convolutional code with generator polynomials \((1/7/5)\). For fairness in comparison between AF and the other schemes, the transmit power at the relay for AF is scaled by a factor of BE. For simplicity only positive integers are considered here as bandwidth expansion factors (BE = \( K \) denotes a bandwidth expansion factor of \( K \)).

Figure 5 presents the overall communication bit error rate (BER) \( Pr(U(n) \neq U(n)) \) for \( SNR_{SR} = 3 \text{ dB} \) and \( SNR_{RD} = 5.1 \text{ dB} \). This scenario is an example showing the conditions described in II-C. Two realizations\(^1\)of our system appear in the plot implementing 1-bit and 2-bit SQ respectively. That is, no bandwidth expansion is required in the first case (BE = 1) while double bandwidth is required for the second one (BE = 2). The schemes are optimized for \( SNR_{SD} = 1 \text{ dB} \) and \( SNR_{SD} = 2.5 \text{ dB} \) respectively. As a reference we have also included the performance of a non-cooperative system (decoding based on single reception of \( Y_{SD} \)) and that of a DAS scheme (decoding of optimally combined \( Y_{SR} \) and \( Y_{SD} \)).

From this figure we can see that our system outperforms AF for a fixed BE factor. Moreover, our system requires less bandwidth to achieve the same performance than AF. That is, roughly we have \( BE_{CF} \approx BE_{AF} + 1 \). We also see that our proposed CF scheme for BE = 1 approaches the best achievable performance for the systems presented in [5] and [6]. This is the performance obtained obtained when decoding \( Y_{SD} \) using the LLRs on the information bits obtained from \( Y_{SR} \) as a priori information. To achieve this performance these systems would require \( C_{RD} \rightarrow \infty \). For a bandwidth expansion factor of BE = 2 we see that our system yields a gain of 1 dB over the CF schemes from [5] and [6] and around 0.75 dB

\(^1\)Only resolution-constrained SQs with \( 2^p \) (\( p \in N^+ \)) reconstruction points are considered here. As a consequence there is a minimum \( SNR_{RD} \) required to allow for iterative decoding. This can be overcome using entropy-constrained SQ in combination with JSCC. However, due to page constraints this is omitted here.
over AF with the same BE. This is due to the fact that for low and medium values of the SNR on the RD link, analog techniques perform worse than digital ones.

Figure 6 compares the effect of bandwidth increase on our proposed scheme and on other alternatives. The scenario depicted here corresponds to cooperation between two users with observations of the same quality (\(\text{SNR}_{SR} = \text{SNR}_{SD} = 3 \, \text{dB}\)). In this case the RD link has \(\text{SNR}_{RD} = 5.5 \, \text{dB}\) and without cooperation we have \(\text{BER} = 4.4 \cdot 10^{-3}\). For the size-constrained SQs considered here this is only optimal for BE = 3. This means that the gap to AF is even larger under the appropriate channel conditions.

![Fig. 5. BER performance comparison. Fig. 6. Effect of bandwidth expansion](image)

V. CONCLUSION

In this paper we have presented a new practical implementation of CF. Our proposed scheme differs from previously proposed ones in the signal that is to be compressed. We propose to quantize directly the observation from the source symbols instead of quantizing a measure on the information bits. Moreover, our implementation uses the redundancy left in the quantized signal as protection against channel errors in the relay-to-destination channel rather than eliminating it. To recover the signal conveyed from the relay, the destination requires from iterative decoding using its direct observation as side information. However, due to the presence of elements that introduce memory at both source and relay, optimal decoding becomes unfeasible. We propose a suboptimal decoding algorithm that disregards partially the statistical dependencies in the data due to these memories and allows for cooperation without major performance losses.

Our system is designed to allow for user cooperation in the cases when neither DF nor AF fully exploit the presence of a relay node. That is, the system exhibits its best performance when the quality of both source-to-relay and relay-to-destination channels is low or moderate. Under these circumstances the combination of an algorithm which does not require from decoding at the relay and the use of a digital modulation outperforms the above stated alternatives. Furthermore, our system in combination with bandwidth expansion in the second phase of the transmission improves communication, reducing the error rates to those of distributed antenna systems.

REFERENCES