KTH Electrical Engineering

Optimized Low-Delay Source–Channel–Relay Mappings

JOHANNES KARLSSON AND MIKAEL SKOGLUND

Stockholm 2010

School of Electrical Engineering and the ACCESS Linnaeus Center,
Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden

IR-EE-KT 2010:023

Optimized Low-Delay Source–Channel–Relay Mappings

Johannes Karlsson, *Student Member, IEEE,* and Mikael Skoglund, *Senior Member, IEEE*

***Abstract*—The three-node relay channel with a Gaussian source is studied for transmission subject to a low-delay constraint. A design algorithm for joint source–channel mappings is proposed and numerically evaluated. The designed system is compared with reference systems, based on modular source and channel coding, and the distortion-rate function for the Gaussian source using known achievable rates for the relay channel. There is a significant gain, in terms of decreased power, in using the (locally) optimized systems compared with the reference systems. The structure of the resulting source mapping and the relay mapping is visualized and discussed in order to gain understanding of fundamental properties of optimized systems. Interestingly, the design algorithm generally produces relay mappings with a structure that resembles Wyner–Ziv compression.**

***Index Terms*—Estimation, joint source–channel coding, relay channel, quantization, sensor networks.**

I. Introduction

The relay channel has been studied extensively since its introduction [1]. With the increasing popularity and relevance of ad-hoc wireless sensor networks, cooperative transmission is more relevant than ever. In this paper, we focus on relaying in the context of source transmission over a sensor network. A sensor node encodes measurements and communicates these to a sink node, with another node acting as a relay in the transmission. We focus on low-delay memoryless source–channel and relay mappings, subject to power constraints at the source and relay nodes. Hence, the proposed technique is a suitable candidate in applications with strict delay and energy constraints, such as in wireless sensor networking for closed-loop control over wireless channels [2], [3].

Existing work on source and channel coding over the relay channel includes [4], [5]. However, whereas [4] looks at asymptotic high-SNR properties the present work is design oriented. Also, although [5] includes some practical results it relies on powerful channel codes. Because of this, the decoding is not instantaneous but a significant delay is needed for the message to be decoded. Another recent study is the one presented in [6]. This work also focuses on characterizing the achievable high-SNR performance, however, in the presence of partial channel-state feedback.

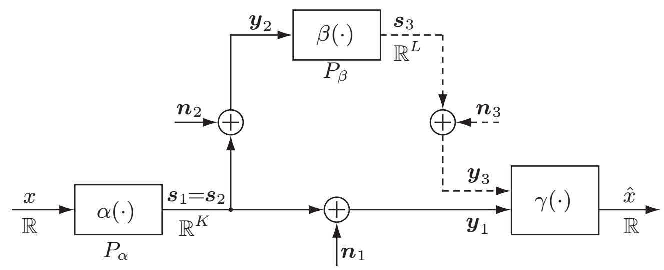

Fig. 1. Structure of the system.

The source–channel separation theorem [7] states that source and channel coding can be treated separately. However, in the case of *low-delay* constraints this is no longer true. We therefore propose a joint source–channel coding solution where at the source node, the source and channel codes are merged into one single operation — a mapping from the source space to the channel space. In a similar way, the operation at the relay is a mapping from its input channel space to its output channel space. We investigate how to optimize[1] both the source–channel mapping at the source as well as the channel–channel mapping at the relay. To our knowledge, there are no similar existing results in this direction. Our approach is however related to the ones being used for bandwidth compression–expansion in [8]–[10] and distributed source–channel coding in [11].

II. Problem Formulation

We will study the three-node system depicted in Figure 1. Our goal is to transmit information about the Gaussian random variable X with variance $\sigma_X^2 = 1$ from the source node to the destination node so that it can be reconstructed with the smallest possible distortion. Besides the direct link we also have a path from the source to the destination via the relay node. The rules for the communication are the following. For each source sample X we have T channel uses at hand. The source and the relay do not transmit at the same time but must share these channel uses. We therefore use K channel uses for the transmission from the source and the remaining $L = T - K$ channel uses for the transmission from the relay. The scenario is in other words that of a half-duplex orthogonal relay channel. All transmissions are disturbed by

Paper approved by J. Kliewer, the Editor for Iterative Methods and Cross-Layer Design of the IEEE Communications Society. Manuscript received November 14, 2008; revised June 15, 2009 and September 21, 2009.

The authors are with the School of Electrical Engineering and the ACCESS Linnaeus Center, at the Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden. (e-mail: {johk, skoglund}@ee.kth.se).

This work was funded in part by the Swedish governmental agency for innovation systems (VINNOVA) and by the European Commission Fp7 project FeedNetBack.

Parts of this paper were presented at IEEE ICASSP 2008.

Digital Object Identifier 10.1109/TCOMM.2010.080595

[1]We use the term *optimized*, in contrast to *optimal*, to refer to a system which is locally optimal but not necessarily globally optimal.

additive white Gaussian noise; the received symbols on each channel can therefore be expressed as

$$\boldsymbol{y}_i = \boldsymbol{s}_i + \boldsymbol{n}_i \quad i = 1, 2, 3, \tag{1}$$

where $\boldsymbol{s}_i$ is the transmitted symbol and $\boldsymbol{n}_i$ is independent white Gaussian noise with $E[\boldsymbol{n}_i \boldsymbol{n}_i^T] = \sigma_i^2 I,\ i = 1, 2, 3.$ The transmitted symbols are given by the functions α and β according to

$$\boldsymbol{s}_1 = \boldsymbol{s}_2 = \alpha(x) \in \mathbb{R}^K, \tag{2}$$

$$\boldsymbol{s}_3 = \beta(\boldsymbol{y}_2) \in \mathbb{R}^L. \tag{3}$$

The equality $\boldsymbol{s}_1 = \boldsymbol{s}_2$ is due to the broadcast nature of a wireless channel. The source and the relay node operate under average transmit power constraints given by

$$\frac{1}{K} E[\|\alpha(X)\|^2] \le P_\alpha, \tag{4}$$

$$\frac{1}{L} E[\|\beta(\boldsymbol{Y}_2)\|^2] \le P_\beta. \tag{5}$$

For notational convenience we define the channel power gain of each channel as $a_i = 1/\sigma_i^2$. Assuming (4) is fulfilled with equality, the total signal-to-noise ratio (SNR) of the transmission from the source node to the destination node is hence given by $P_\alpha a_1$. We further assume that the total SNR of all channels is known by all involved parts. The destination node receives two symbols — $\boldsymbol{y}_1$ from the direct link and $\boldsymbol{y}_3$ from the relay. Based on these the transmitted value is estimated as

$$\hat{x} = \gamma(\boldsymbol{y}_1, \boldsymbol{y}_3). \tag{6}$$

Given this system we want to find the optimal source mapping, relay mapping, and receiver — denoted α, β, and γ. To have a low-delay system we want the source and the relay nodes to work on a sample-by-sample basis restricting K and L to be integers. If $K > 1$, α will in general be a nonlinear mapping from the one-dimensional source space to the K-dimensional channel space. In a similar way β will be a nonlinear mapping from the K-dimensional input of the relay to its L-dimensional output. As distortion measure we use the mean squared error (MSE), $E[(X - \hat{X})^2]$, "optimal" therefore refers to optimal in the minimum MSE (MMSE) sense.

III. Optimized Mappings

The expected distortion for a given system can be written as

$$D = E[(X - \hat{X})^2] = \iiiint p(x) p(\boldsymbol{y}_1|\alpha(x)) p(\boldsymbol{y}_2|\alpha(x)) \times p(\boldsymbol{y}_3|\beta(\boldsymbol{y}_2))(x - \gamma(\boldsymbol{y}_1, \boldsymbol{y}_3))^2 \mathrm{d}x \mathrm{d}\boldsymbol{y}_1 \mathrm{d}\boldsymbol{y}_2 \mathrm{d}\boldsymbol{y}_3, \tag{7}$$

where $p(\cdot)$ and $p(\cdot|\cdot)$ denote probability density functions (pdfs) and conditional pdfs, respectively. The factorization of $p(x, \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$ in (7) follows from the fact that all channels are orthogonal with independent noises. What we would like is to find α, β, and γ such that D is minimized given the power constraints in (4) and (5). There are two problems with this direct approach. First, it is very hard to optimize all parts of the system simultaneously; second, the optimal mappings could be arbitrary nonlinear mappings with no closed form expressions. To make the problem feasible we take the following suboptimal approach. Instead of optimizing all parts of the system simultaneously we use the common strategy of optimizing one part at a time keeping the others fixed. The second problem is solved by discretizing each dimension of the channel space into M equally spaced points with spacing Δ according to

$$\mathcal{S} = \{-\Delta\frac{M-1}{2}, -\Delta\frac{M-3}{2}, \ldots, \Delta\frac{M-3}{2}, \Delta\frac{M-1}{2}\} \tag{8}$$

and restricting the outputs of the source and the relay node to satisfy $\boldsymbol{s}_1 \in \mathcal{S}^K$ and $\boldsymbol{s}_3 \in \mathcal{S}^L$, respectively. At the receiving side the same approximation is made using a hard decision decoding rule — for instance, $\boldsymbol{y}_1$ is decoded according to

$$\hat{\boldsymbol{y}}_1 = \arg \min_{\boldsymbol{y}'_1 \in \mathcal{S}^K} \|\boldsymbol{y}_1 - \boldsymbol{y}'_1\|^2, \tag{9}$$

where "ˆ" will be used to indicate that the value has been discretized. This approximation is expected to be good as long as M is sufficiently large and Δ is small in relation to the standard deviation of the channel noise, σ_i. In the following analysis $P(\cdot|\cdot)$ will be used for conditional probabilities — for example, $P(\hat{\boldsymbol{y}}_3|\boldsymbol{s_1})$ denotes the probability that the relay receives $\hat{\boldsymbol{y}}_3$ given that $\boldsymbol{s}_1$ is transmitted from the source.

A. Optimal Source Mapping

The problem of finding the optimal source mapping α (assuming β and γ are fixed) is a constrained optimization problem, which can be turned into the following unconstrained problem using the Lagrange multiplier method [12], [13]

$$\min_\alpha \left(E[(X - \hat{X})^2] + \lambda E[\|\alpha(X)\|^2] \right), \tag{10}$$

where

$$E[(X - \hat{X})^2] = \int p(x) E[(x - \hat{X})^2|\alpha(x)] \mathrm{d}x, \tag{11}$$

$$E[\|\alpha(X)\|^2] = \int p(x) \|\alpha(x)\|^2 \mathrm{d}x. \tag{12}$$

Since $p(x)$ in (11)–(12) is nonnegative, it is clear that the operation of the source mapping, α, can be optimized for each x individually according to

$$\alpha(x) = \arg \min_{\boldsymbol{s}_1 \in \mathcal{S}^K} \left(E[(x - \hat{X})^2|\boldsymbol{s}_1] + \lambda \|\boldsymbol{s}_1\|^2 \right) \tag{13}$$

where

$$E[(x - \hat{X})^2|\boldsymbol{s}_1] = \sum_{\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2, \hat{\boldsymbol{y}}_3} P(\hat{\boldsymbol{y}}_1|\boldsymbol{s}_1) P(\hat{\boldsymbol{y}}_2|\boldsymbol{s}_1) \times P(\hat{\boldsymbol{y}}_3|\beta(\hat{\boldsymbol{y}}_2))(x - \gamma(\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_3))^2. \tag{14}$$

The intuition behind the Lagrange term $\lambda\|\boldsymbol{s}_1\|^2$ is the following: $\|\boldsymbol{s}_1\|^2$ is a measure of the power that is needed to transmit the signal $\boldsymbol{s}_1$, the term $\lambda\|\boldsymbol{s}_1\|^2$ can therefore be used to control the transmit power of the source node by penalizing signals that would use too much power. When $\lambda \ge 0$ is set to the "correct" value, the source node will not map x to the signal that gives the lowest distortion but rather to the signal that gives the lowest distortion conditioned that the power constraint in (4) is fulfilled.

B. Optimal Relay Mapping

In a similar way, the minimization to find the optimal relay mapping β (assuming α and γ are fixed), can be turned into the following unconstrained minimization problem

$$\min_{\beta}\left(E[(X-\hat{X})^2]+\eta E[\|\beta(\hat{\boldsymbol{Y}}_2)\|^2]\right), \quad (15)$$

where

$$E[(X-\hat{X})^2]=\sum_{\hat{\boldsymbol{y}}_2}P(\hat{\boldsymbol{y}}_2)E[(X-\hat{X})^2|\hat{\boldsymbol{y}}_2,\beta(\hat{\boldsymbol{y}}_2)], \quad (16)$$

$$E[\|\beta(\hat{\boldsymbol{Y}}_2)\|^2]=\sum_{\hat{\boldsymbol{y}}_2}P(\hat{\boldsymbol{y}}_2)\|\beta(\hat{\boldsymbol{y}}_2)\|^2. \quad (17)$$

Equations (11) and (16) are two different ways of expanding the MSE using Bayes' rule. Looking at (16) and (17), it is once again clear that the minimization can be done individually for each $\hat{\boldsymbol{y}}_2\in\mathcal{S}^K$, which gives

$$\beta(\hat{\boldsymbol{y}}_2)=\arg\min_{\boldsymbol{s}_3\in\mathcal{S}^L}\left(E[(X-\hat{X})^2|\hat{\boldsymbol{y}}_2,\boldsymbol{s}_3]+\eta\|\boldsymbol{s}_3\|^2\right) \quad (18)$$

where

$$E[(X-\hat{X})^2|\hat{\boldsymbol{y}}_2,\boldsymbol{s}_3]=\textstyle\sum_{\hat{\boldsymbol{y}}_1,\hat{\boldsymbol{y}}_3}P(\hat{\boldsymbol{y}}_3|\boldsymbol{s}_3)\times \int_x p(x|\hat{\boldsymbol{y}}_2)P(\hat{\boldsymbol{y}}_1|\alpha(x))(x-\gamma(\hat{\boldsymbol{y}}_1,\hat{\boldsymbol{y}}_3))^2\mathrm{d}x. \quad (19)$$

In (18), $\eta\geq 0$ is the Lagrange multiplier which — when chosen correctly — makes sure that the power constraint (5) is satisfied.

Sawtooth Mappings ($K=L=1$)**:** As we will see in Section IV, all of the optimized relay mappings have a similar shape in the one-dimensional case (i.e., $K=L=1$). Based on this observation we propose to use a sawtooth mapping as shown in Figure 2. This mapping has previously been proposed for distributed source–channel coding [14] and also for the relay channel in the context of maximum achievable rates [15].

The sawtooth mapping can be parametrized by the two parameters b and c and is defined as

$$\beta(y_2)=\begin{cases} cy_2 & \text{if } y_2\in[-b,b)\\ \beta(y_2-2bm) & \text{if } y_2-2bm\in[-b,b),\ m\in\mathbb{Z},\end{cases} \quad (20)$$

where, for a given b, the parameter c must be chosen so that the power constraint in (5) is satisfied, that is, $E[\beta^2(Y_2)]=P_\beta$. The optimal value of b will depend on the channel gains and is easiest found by performing a grid search.

C. Optimal Receiver

Since we use the MSE as a distortion measure, it is a well known fact from estimation theory that the optimal receiver (assuming α and β are fixed) is the expected value of X given the received symbols,

$$\hat{x}=\gamma(\hat{\boldsymbol{y}}_1,\hat{\boldsymbol{y}}_3)=E[X|\hat{\boldsymbol{y}}_1,\hat{\boldsymbol{y}}_3]=\int_x xp(x)\times \sum_{\hat{\boldsymbol{y}}_1}P(\hat{\boldsymbol{y}}_1|\alpha(x))\sum_{\hat{\boldsymbol{y}}_2}P(\hat{\boldsymbol{y}}_2|\alpha(x))\sum_{\hat{\boldsymbol{y}}_3}P(\hat{\boldsymbol{y}}_3|\beta(\hat{\boldsymbol{y}}_2))\mathrm{d}x. \quad (21)$$

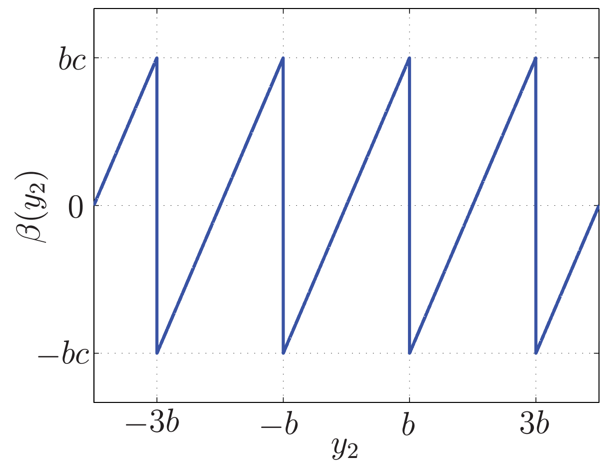

Fig. 2. Parameterized sawtooth mapping.

As an alternative receiver for the sawtooth mappings, we will also implement the maximum likelihood (ML) decoder given by

$$\hat{x}=\gamma_{\mathrm{ML}}(y_1,y_3)=\arg\max_x p(y_1,y_3|x). \quad (22)$$

The ML decoder is suboptimal in the sense that it does not minimize the MSE.

D. Design Algorithm

Given the above expressions for the source mapping, the relay mapping, and the receiver it will be possible to optimize the system iteratively. We do this by keeping two parts of the system fixed while we optimize the third part. One common problem with an iterative technique like the one suggested here is that the final solution will depend on the initialization of the algorithm, if the initialization is bad we are likely to end up in a poor local minimum. One method that has proven to be helpful in counteracting this is noisy channel relaxation [9], [16] which works in the following way. A system is first designed for a noisy channel, the solution obtained is then used as an initialization when designing a system for a less noisy channel. The noise is reduced and the process is repeated until the desired noise level is reached. The intuition behind this method is that an optimal system for a noisy channel has a simple structure and is easy to find, as the channel noise is decreased more structure is gradually added to form the final system. Given a scenario where K and L are specified, the design algorithm is formally stated below.

1) Choose some initial mappings for β and γ.
2) Let $\boldsymbol{A}=(a_1,a_2,a_3)$ be the channel power gains for which the system should be optimized. Create $\boldsymbol{A'}=(a_1',a_2',a_3')$, where $a_1'\leq a_1, a_2'\leq a_2, a_3'\leq a_3$ (i.e., $\boldsymbol{A'}$ corresponds to a channel which is more noisy than $\boldsymbol{A}$).
3) Design a system for $\boldsymbol{A'}$ according to:
   a) Set the iteration index $k=0$ and $D^{(0)}=\infty$.
   b) Set $k=k+1$.
   c) Find the optimal source mapping α by using (13).
   d) Find the optimal receiver γ by using (21).
   e) Find the optimal relay mapping β by using (18).

f) Find the optimal receiver γ by using (21).

g) Evaluate the distortion $D^{(k)}$ for the system. If the relative improvement of $D^{(k)}$ compared to $D^{(k-1)}$ is less than some threshold $\delta > 0$ go to Step 4. Otherwise go to Step (b).

4) If $\boldsymbol{A}' = \boldsymbol{A}$ stop the iteration. Otherwise increase $\boldsymbol{A}'$ according to some scheme (e.g., linearly) and go to Step 3 using the current system as initialization when designing the new system.

IV. Simulation Results

To evaluate the algorithm we have designed systems for different combinations of K and L. We will compare the performance against some reference systems, given below, and the distortion-rate function for a memoryless Gaussian source [7] using the achievable rate of the compress-and-forward (CF) scheme [17] (assuming orthogonal transmissions).

A. Reference Systems

$\boldsymbol{K = L = 1}$ **:** For the one-dimensional case we use linear transmission at the source node in conjunction with estimate-and-forward (EF) at the relay as our reference system. For EF, the relay function β is given by $\beta(y_2) = cE[s_2|y_2]$. It should be noted that in the case of a Gaussian source and linear transmission at the source node, amplify-and-forward is equivalent to estimate-and-forward.

$\boldsymbol{K = 2, L = 1}$ **:** In this case, we compare our optimized system with two different reference systems. The first system operates by transmitting the source sample X directly on the channel for both channel uses (scaled to fulfill the power constraint), that is, repetition coding, and uses EF at the relay. This system will be denoted *Linear*. One disadvantage of this scheme is the repetition coding in the transmission from the source node. To better fill the two-dimensional channel space we propose the following alternative system, denoted *Digital*, where we have taken off-the-shelf components and put them together in a modular fashion. Instead of the source mapping $\alpha(\cdot)$ we use a 16-level Lloyd–Max quantizer [18], [19] followed by a 16-QAM mapping to the channel space. The relay node makes a hard decision on the received signal and modulates the decoded symbol with 16-PAM. At the destination node the received signals are once again decoded with a hard decision and finally x is reconstructed as the expected value of x given the decoded symbols. This system is optimized in the sense that we use a source-optimized quantizer, a good choice of the mapping to QAM symbols (i.e., a mapping that corresponds to a good *index assignment*, so that neighboring quantization levels correspond to neighboring QAM symbols [20]), and an optimal receiver (given the hard decoded received symbols).

$\boldsymbol{K = 1, L = 2}$ **:** As in the one-dimensional case, we use linear transmission at the source node and study two different relay mappings — a linear repetition code and a digital system, denoted *Linear* and *Digital*, respectively. The Linear relay mapping scales the input to satisfy the power constraint and transmits the same symbol two times. The source symbol, x, is then estimated as the expected value given the received signals. The Digital system performs a 16-level quantization (optimized for the input distribution) and transmits the quantization index using 16-QAM. At the receiver, the quantization index is decoded using a hard decision ML-decoding rule. Finally, the hard decoded index is used in conjunction with the value received on the direct link to find the expected value of the source symbol given these values.

B. Implementation Aspects

In Step 1 of the design algorithm, β was initialized as a linear mapping and γ was randomly initialized. However, it is important to understand that the use of noisy channel relaxation makes the solution less sensitive to the initialization. In the case of the relay channel, with three different channels, the problem is instead that of choosing a starting point and a path for the noisy channel relaxation. For the case $K = L = 1$, we started at $\boldsymbol{A}_1' = (a_1, -5, -5)$ dB and linearly increased the second and third components one at a time until they reached their corresponding final values. For the other two cases, we started at $\boldsymbol{A}_2' = (-5, -5, -5)$ dB and linearly increased all components simultaneously until they reached $\boldsymbol{A}$. To reduce the complexity of the design algorithm in the case $K = 1$, we fixed α to be a linear scaling (fulfilling the power constraint) followed by a mapping to the closest point in the set $\mathcal{S}$. Steps 3c) and 3d) were omitted in the design algorithm for these systems. Although there are no proofs that this is the jointly optimal strategy, it can be justified by the fact that linear scaling is individually optimal for each point-to-point link from the source node in the case $K = 1$. A final note regarding the Lagrange multipliers λ and η. After each iteration in the design algorithm, they were either increased or decreased in small steps depending on whether the used power was too high or too low.

Another important aspect is the number of points, M, in the discretization of the channel space given by (8). In our implementation we have varied M with the total channel SNR, using a lower resolution for low SNRs and a higher resolution for high SNRs. For example, at an SNR of 5 dB we have used $M = 64$ and at 25 dB we have used $M = 512$. Δ has been varied along with M according to $\Delta = 8/(M-1)$; meaning that we have a good approximation of the channel in the interval $[-4, 4]$. There is a tradeoff in the choice of M and Δ, increasing M increases the complexity of the design algorithm. If on the other hand M is too small, the distortion created by the discrete approximation is significant.

Finally, we will give some details of the actual implementation of (13), (18), and (21). All integrals with respect to x and also the source mapping α, have been calculated using a set of training samples which turns the integrals into sums. The size of this set has been 200000 in the case of a $K = 1$ and 10000 in the case of $K = 2$. Since the channel space is approximated by the finite set $\mathcal{S}$, the relay mapping and the receiver can be stored as lookup tables.

C. Numerical Results

In the following simulations, we assume that the source mapping and relay mapping are optimized for certain signal-to-noise ratios (SNRs), marked with circles in the figures, but that the receiver has perfect channel state information

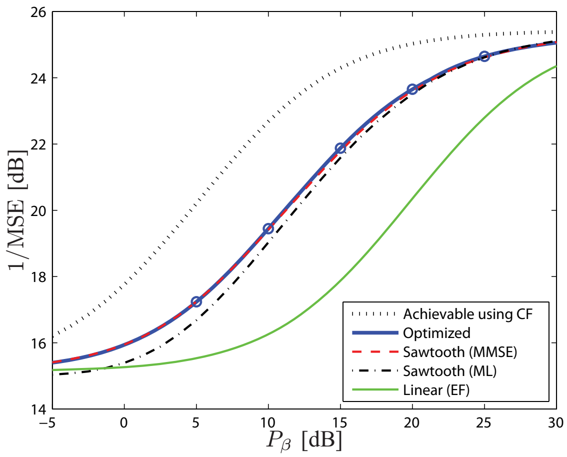

Fig. 3. $K = L = 1$ Simulation results when P_β is varied while $P_\alpha = 0$ dB and $a_1 = 15$ dB, $a_2 = 25$ dB, and $a_3 = 0$ dB. The circles mark the points for which the system is optimized.

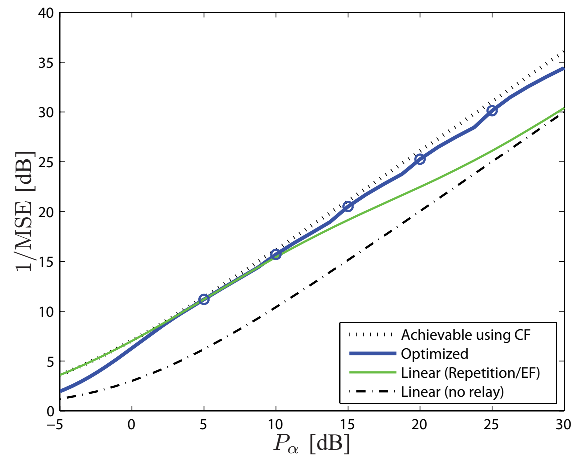

Fig. 4. $K = L = 1$ Simulation results when P_α is varied while $P_\beta = 0$ dB and $a_1 = 0$ dB, $a_2 = 5$ dB, and $a_3 = 20$ dB. The circles mark the points for which the system is optimized.

and therefore adapts to the current channel state using (21). We will mainly study the power efficiency of the relay node, that is, how much power the relay needs to achieve a certain performance. For the one-dimensional case, which we study more extensively, we have also included results showing the power efficiency of the source node for different relay mappings.

$\boldsymbol{K = L = 1}$ **:** If the quality of the link to the relay is better than the direct link, as in Figure 3, the relay can improve the performance significantly. The horizontal power gain[2] of using the optimized system over the linear system is as much as 7–8 dB in the entire region shown. It should be noted that this increase is only due to utilizing the power in a more efficient way and comes at virtually no extra complexity in the relay. The gap to the achievable rate is quite significant, around 6.5–8 dB for the optimized points. This gap will be discussed later on. It is also evident that the optimized mappings and the sawtooth mappings with MMSE receiver (given by (21)) perform almost the same (the optimized mappings are about 0.1 dB better than the sawtooth mappings at the design points), making them practically impossible to distinguish. It also turns out that the ML detector (given by (22)) performs very close to the optimal MMSE detector, which is encouraging due to its simplicity. It should be emphasized that the sawtooth mappings have been optimized for each SNR point and each detector. A sawtooth mapping which is optimal for the MMSE detector is not necessarily optimal for the ML detector. In Figure 4, we vary the power of the source node. In this case the optimized system manages to follow the achievable curve closely — the gap is only 0.1 dB at $P_\alpha = 5$ dB and increases slightly with the SNR to 0.7 dB at $P_\alpha = 25$ dB. This can be explained as follows, up to some point, say $P_\alpha = 10$ dB, the channel from the relay to the destination is much better than the channels from the source. This implies that all relay mappings perform basically the same as long as they are non-destructive and do not discard any information (c.f. $P_\beta \to \inf$). For this reason, also the linear mapping performs very well.

[2]We will only consider this gain.

The fact that we are close to the achievable curve strengthens our previous intuitive suggestion that linear transmission at the source node works well for $K = 1$. As the power of the source node increases further, we see that the linear relay mapping approaches the same performance as not using the relay at all. The relatively high noise power on the channel from the relay to the destination makes the information from a linear relay unusable. It is therefore interesting to note how well the optimized mappings follow the achievable curve. As the power of the source node increases, the correlation between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ will also increase. The optimized mappings take advantage of this increasing correlation and perform a kind of Wyner–Ziv compression where $\boldsymbol{y}_2$ is used as side information when decoding the information from the relay. An example of how this is done will be given in Section IV-D.

$\boldsymbol{K = 2, L = 1}$ **:** In this case (Figure 5), we have the additional problem of designing a good source mapping, α. The optimized system still has a significant gain over the linear system, ranging from 5 dB at $P_\beta = 5$ dB to 10 dB at $P_\beta = 15$ dB. The digital system performs slightly worse than the linear system. From the figure, the different systems does not seem to reach the same performance as the power of the relay increases. This is in fact true, the achievable curve reaches a limit of 31 dB whereas the performance of the linear system is limited to 18.5 dB. This gap is due to the linear system's inability to produce a two-dimensional distribution that matches the Gaussian channel from the source. The source mapping used in the optimized systems (see Section IV-D) does a better job, but does clearly not achieve the capacity on the two-dimensional channel from the source node either. Similar results for bandwidth expansion curves can be observed in [10].

$\boldsymbol{K = 1, L = 2}$ **:** Changing the situation, having one channel use for the source transmission and two channel uses for the relay transmission, the results are similar to the one-dimensional case as can be seen in Figure 6. The gap to the linear system is around 3 dB and the gap to the achievable curve is ranging from 4.5 dB at $P_\beta = 5$ dB to 8 dB at $P_\beta = 15$

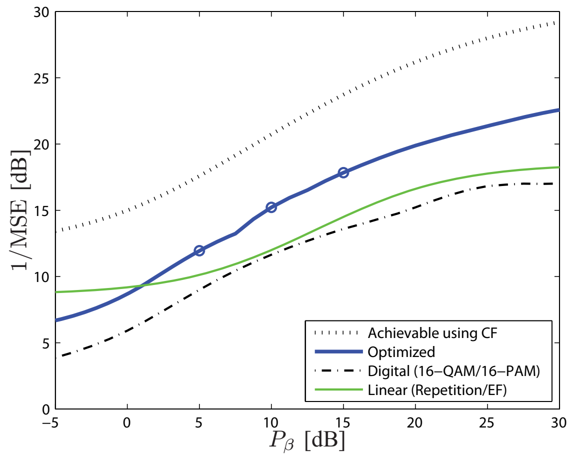

Fig. 5. $K = 2, L = 1$ Simulation results when P_β is varied while $P_\alpha = 0$ dB and $a_1 = 5$ dB, $a_2 = 15$ dB, and $a_3 = 0$ dB. The circles mark the points for which the system is optimized.

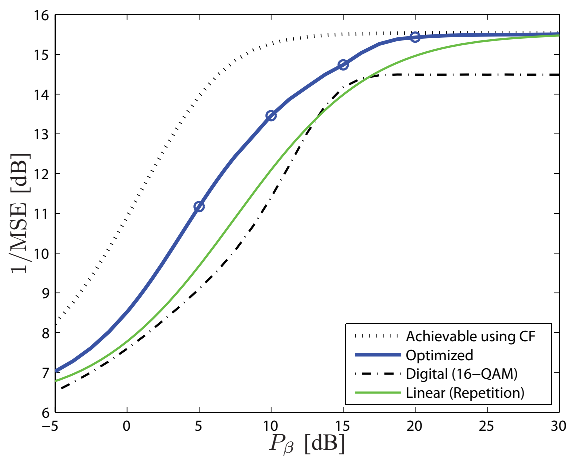

Fig. 6. $K = 1, L = 2$ Simulation results when P_β is varied while $P_\alpha = 0$ dB and $a_1 = 5$ dB, $a_2 = 15$ dB, and $a_3 = 0$ dB. The circles mark the points for which the system is optimized.

dB.

The significant gap to the achievable curve in most cases can to a large extent be explained by our low-delay one-dimensional approach where we transmit one sample at a time, in contrast to the infinite dimensions used in the proofs for both the distortion-rate function and the achievable rate. An exception to this is when there is no side information available and the distribution of the source matches the channel, in which case uncoded transmission is optimal (e.g., transmitting a one-dimensional Gaussian variable on a Gaussian channel).

D. Structure of β

$\boldsymbol{K = L = 1}$ **:** Figure 7 shows an example of a typical relay mapping in the one-dimensional case. It is clear that the proposal of sawtooth mappings in Section III-B is well motivated. The main reason why this optimized mapping performs better than a linear mapping is the steeper slope, which effectively decreases the impact of the channel noise. Looking at the sawtooth mapping in Figure 2, one could say

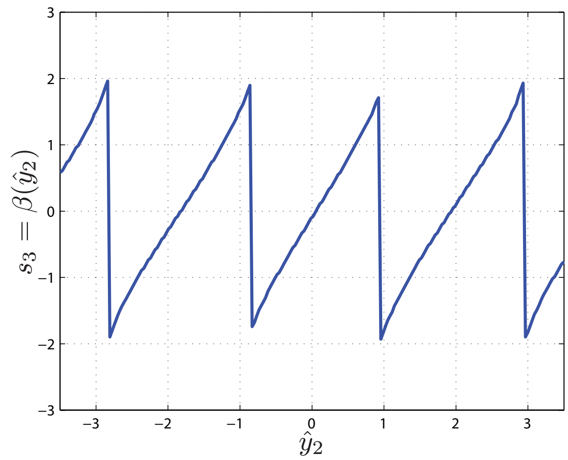

Fig. 7. Relay mapping ($K = L = 1$) optimized for $P_\alpha = P_\beta = 0$ dB and $a_1 = 15$ dB, $a_2 = 15$ dB, and $a_3 = 20$ dB.

that decreasing b allows us to increase c (without violating the power constraint) and therefore get lower distortion. It is apparent that the relay mapping is not injective since several input values are mapped to the same output value; this way of reusing output values can be seen as *Wyner–Ziv* compression[3]. The reuse of output values is only possible due to the side information from the direct link. Returning to Figure 2 and assuming that the output of the relay is 0, in this case the side information will provide the necessary information to determine whether y_2 was, for example, $-2b$, 0, or $2b$. However, if b is decreased below a certain threshold (dependent on a_1, a_2, and a_3), the probability of making the wrong decision based on the side information will be significant and the decoder will therefore make large estimation errors. It is in particular the values near the discontinuities that are sensitive to large estimation errors. Looking at the optimized mapping again, one can see that the slope is slightly steeper near the discontinuities. The extra energy spent for these values increases the distance between points in the safe region (far away from the discontinuities) and the critical points (near the discontinuities). This could be the explanation of the slightly better performance of the optimized mappings compared with that of the sawtooth mappings. It is quite remarkable that the design algorithm produces the sawtooth-like mappings despite the fact that the initial relay mapping is *linear*. We believe that this is a consequence of the channel relaxation, especially the fact that a_3 is the last component that is increased, and the Lagrange multipliers and how these are updated in small steps.

$\boldsymbol{K = 2, L = 1}$ **:** The source mapping α is now a mapping from the one-dimensional source space to the two-dimensional channel space. An example of such a mapping is shown in the left part of Figure 8, where the curve shows how input symbols in the interval $[-3, 3]$ are mapped to two-dimensional output symbols. The mapping is such that small negative values of x are mapped to one end of the curve and as x is increased

[3]The Wyner-Ziv scheme saves *rate* by sending an ambiguous "bin-index" rather than a codeword index. The ambiguity is resolved at the decoder by using the side information to identify the correct codeword in the bin. In our case, the relay saves *power* by informing the receiver about a set (a "bin") of possible values, rather than a specific value.

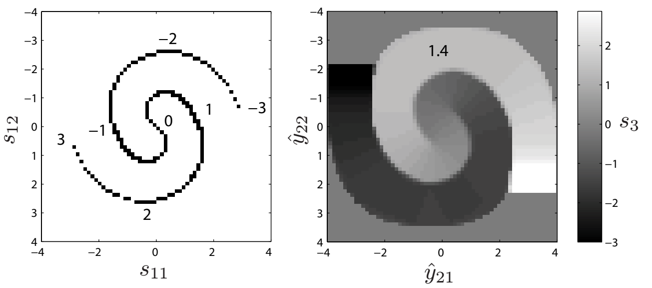

Fig. 8. Structure of α (to the left) and β (to the right) ($K = 2, L = 1$) optimized for $P_\alpha = P_\beta = 0$ dB and $a_1 = 5$ dB, $a_2 = 15$ dB, and $a_3 = 10$ dB. In the left part, it is shown how the interval $[-3, 3]$ of the one-dimensional input is mapped to the two-dimensional output $\boldsymbol{s}_1 = (s_{11}, s_{12})$. In the right part, the color in the figure together with the colorbar shows how the two-dimensional input, $\hat{\boldsymbol{y}}_2 = (\hat{y}_{21}, \hat{y}_{22})$, is mapped to the one-dimensional output s_3.

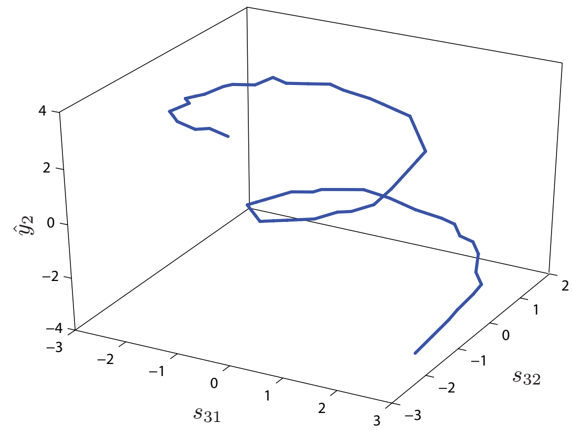

Fig. 9. Relay mapping ($K = 1, L = 2$) optimized for $P_\alpha = P_\beta = 0$ dB and $a_1 = 5$ dB, $a_2 = 15$ dB, and $a_3 = 5$ dB. The two-dimensional output, shown on the x- and y-axes, as a function of the one-dimensional input, shown on the z-axis. In other words $\boldsymbol{s}_3 = \beta(\hat{y}_2)$, where $\boldsymbol{s}_3 = (s_{31}, s_{32})$.

the mapping follows the curve to the other end. Values around zero — which are the most likely values for a Gaussian source — are mapped to the center of the curve which lies close to the origin where $\|\boldsymbol{s}_1\|^2$ is small. The transmission power for these values is hence minimized. In contrast, values that are less probable are instead mapped to points in the channel space that use more energy. This structure is due to the Lagrange term in (18); similar results have been been obtained in [9]–[11]. Due to the high noise level on the direct link, the destination cannot distinguish between different parts of the curve by only looking at the direct link. For example, the receiver will not be able to determine whether 1 or -3 was transmitted since they are mapped to symbols that are close in the channel space. The relay node needs to help the receiver to distinguish which point, or at least which region, of the curve that was transmitted. Looking at the right part of Figure 8, which shows the relay mapping, we can see that this is exactly what the relay does. Something that is interesting to notice is that the relay is *not* the inverse of the source mapping which it would be if the relay tried to estimate x and send the estimate to the receiver. This is easiest seen by the fact that for some of the outer parts of the curve, the relay uses the same output symbol for large regions (e.g., $s_3 \approx 1.4$ for the upper part of the curve) which means that the relay does not send an estimate of what was received but rather just tells the receiver that the transmitted point was on the upper part of the curve. Using this information the receiver estimates x based on the value received from the direct link conditioned that the transmitted point was on the upper part of the curve.

$\boldsymbol{K = 1, L = 2:}$ In Figure 9, we finally show an example of a mapping where the relay performs an expansion — from its one-dimensional input to its two-dimensional output. Once again, there is a reuse of the output symbols which is only possible due to the side information from the direct link. Looking at the spiral from above, a similarity to the polynomial based source–channel codes proposed in [21] can be seen.

V. Conclusions

We have proposed a low-delay scheme for joint source–channel coding over the relay channel. The design also includes optimizing the relay itself. The numerical results show that the joint design works well and gives better performance than the reference systems. We have also provided useful insight into the structure of the (locally) optimized source–channel and relay mappings, and how these mappings together make it possible for the receiver to output a good estimate of the source. The mapping at the relay reuses output symbols and is clearly reminiscent of Wyner–Ziv compression. Based on observing the structure of our optimized systems, we proposed the use of sawtooth mappings for the case of one-dimensional relaying. The sawtooth mappings can in many cases be used instead of the optimized mappings without any performance degradation.

References

[1] E. C. van der Meulen, “Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, no. 1, pp. 120-154, 1971.
[2] L. Bao, M. Skoglund, and K. H. Johansson, “Iterative encoder-controller design for feedback control over noisy channels," *IEEE Trans. Automatic Control*, accepted for publication.
[3] G. N. Nair, F. Fagnani, S. Zampieri, and R. Evans, “Feedback control under data rate constraints: An overview," in *Proc. IEEE*, vol. 95, no. 1, pp. 108-137, Jan. 2007.
[4] D. Gündüz and E. Erkip, “Source and channel coding for cooperative relaying," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3454-3475, Oct. 2007.
[5] H. Y. Shutoy and D. Gündüz and E. Erkip and Y. Wang, “Cooperative source and channel coding for wireless multimedia communications," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 2, pp. 295-307, Aug. 2007.
[6] T. T. Kim, M. Skoglund, and G. Caire, “On cooperative source transmission with partial rate and power control," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1408-1418, Oct. 2008.
[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.
[8] V. Vaishampayan, “Combined source-channel coding for bandlimited waveform channels," Ph.D. dissertation, University of Maryland, 1989.
[9] A. Fuldseth and T. A. Ramstad, “Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space," in *International Conf. Acoustics, Speech Signal Process. (ICASSP)*, Munich, Germany, Apr. 1997, pp. 3093-3096.
[10] P. A. Floor, T. A. Ramstad, and N. Wernersson, “Power constrained channel optimized vector quantizers used for bandwidth expansion," in *IEEE International Symp. Wireless Commun. Syst.*, Oct. 2007.

[11] N. Wernersson, J. Karlsson, and M. Skoglund, “Distributed quantization over noisy channels," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1693-1700, June 2009.

[12] H. Everett III, “Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, no. 3, pp. 399-417, 1963.

[13] Y. Shoham and A. Gersho, “Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 9, no. 9, pp. 1445-1453, 1988.

[14] N. Wernersson and M. Skoglund, “Nonlinear coding and estimation for correlated data in wireless sensor networks," *IEEE Trans. on Communications*, vol. 57, no. 10, pp. 2932-2939, October 2009.

[15] S. Yao and M. N. Khormuji and M. Skoglund, “Sawtooth relaying," *IEEE Commun. Lett.*, vol. 12, no. 9, Sept. 2008.

[16] S. Gadkari and K. Rose, “Noisy channel relaxation for VQ design," in *International Conf. Acoustics, Speech Signal Process. (ICASSP)*, May 1996, pp. 2048-2051.

[17] A. Høst-Madsen and J. Zhang, “Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020-2040, June 2005.

[18] S. P. Lloyd, “Least Squares Quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129-137, Mar. 1982.

[19] Y. Linde and A. Buzo and R. M. Gray, “An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84-95, Jan. 1980.

[20] M. Skoglund, “On channel-constrained vector quantization and index assignment for discrete memoryless channels," *IEEE Trans. Info. Theory*, vol. 45, no. 7, pp. 2615-2622, Nov. 1999.

[21] N. Wernersson, M. Skoglund, and T. Ramstad, “Polynomial based analog source-channel codes," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2600-2606, Sept. 2009.

Johannes Karlsson (S'07) received the M.Sc. degree in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2006. He joined the Communication Theory Lab, KTH, in 2006, where he is currently working towards the PhD degree. As part of his PhD studies, he received the Tech. Lic. in Telecommunications in January, 2009. His research interests include joint source-channel coding, distributed source coding, and wireless sensor networks.

Mikael Skoglund (S'93-M'97-SM'04) received the Ph.D. degree in 1997 from Chalmers University of Technology, Sweden. In 1997, he joined the Royal Institute of Technology, Stockholm, Sweden, where he was appointed Professor of Communication Theory in October 2003.

Dr. Skoglund's research interests are in information theory, communications, and detection and estimation. He has worked on problems in source–channel coding, coding and transmission for wireless communications, Shannon theory and statistical signal processing. He has authored some 170 scientific papers, including papers that have received awards, invited conference presentations, and papers ranking as highly cited according to the ISI Essential Science Indicators. He has also consulted for industry, and he holds six patents.

Dr. Skoglund has served on multiple technical program committees for IEEE conferences. During 2003–08 he was an associate editor with the IEEE Transactions on Communications and he is presently on the editorial board for IEEE Transactions on Information Theory. He is an IEEE senior member.