

JOINT SOURCE-CHANNEL CODING OF UNIFORM MEMORYLESS SOURCES OVER BINARY SYMMETRIC CHANNELS

Seyed Bahram ZAHIR AZAMI¹, Pierre DUHAMEL² and Olivier RIOUL³

École Nationale Supérieure des Télécommunications
URA CNRS 820

^{1,3}Communications and electronics department

²Signal and image processing department

46, rue Barrault,
75634 Paris Cedex 13,
France

emails: ¹zahir@com.enst.fr, ²duhamel@sig.enst.fr, ³rioul@com.enst.fr

Abstract

A joint design of source and channel coding is considered for digital transmission over a *binary symmetric channel* (BSC).

Three major techniques are used: For a binary source, channel optimized coder design (COCD) is proposed. This method has non-binary decoding values and minimizes the MSE distortion measure. It can be understood as a channel coder specifically designed to be robust under transmission errors. Then, this tool is applied to the transmission of a uniform source. This is done through a bitwise decomposition structure in which the samples of a uniform source are first expressed in binary representation and bits of same weight are encoded together. Finally, a practical optimization algorithm is explained that delivers the best set of coders, for a given *raw error probability* (BER) of the channel.

1 Introduction

This paper addresses the transmission of digital data on noisy channels. The considered sources are *binary symmetric source* (BSS) and uniform source. The transmission is considered to be over a BSC. These choices of source

and channel may seem overly simplistic, however they are studied in the first step, for a better understanding of the problem, and provide later building blocks to be used in more sophisticated systems. Actual communication systems carry sources of various types, which require different error protections. For this reason, we model the channel as a BSC, which encompasses the physical channel as well as the minimum required error protection for any source that is transmitted. The additional protection is source-dependent, and can be merged with the source coder. By adopting this policy, joint source-channel coding can be performed, without specializing the transmission system to a given source.

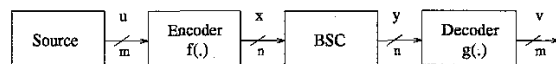


Figure 1: A simple transmission configuration.

A digital transmission configuration is presented in figure 1. The code *rate* is defined as the number of coded bits divided by the number of source bits, $R = \frac{n}{m}$. We seek to minimize a properly defined distortion measure, $D = \frac{1}{m} \mathbf{E}\{\|\mathbf{u} \ominus \mathbf{v}\|^2\}$, with the constraint that the total transmission rate, R , is less than (or

equal) the desired bit rate.

As a result of Shannon's joint source-channel coding theorem, there was a great tendency in the information theory society to do the source and the channel coding separately because they can be treated separately without any loss of performance for the overall system [5]. However, such a tandem source-channel coding necessitates very long blocks and very complex coders. Despite its optimality, Shannon's separation theorem does not necessarily result in the best economic solution.

2 Bitwise decomposition, compared to the bounds

Some bounds In this section, we obtain the relation between the source rate-distortion function and the *optimum performance theoretically attainable* (OPTA), $R(D)$, on a BSC, for a given BER, p . The OPTA function is the expression of the smallest possible distortion as a function of the bit rate, when transmitting a given source on a given channel. According to Shannon's theory [5], the OPTA curve, $R(D)$, is given by:

$$R(D) = \frac{R_s(D)}{C} \quad (1)$$

where $R_s(D)$ is the source rate-distortion function and C is the channel capacity.

For our model, the BSC is parameterized by the BER, p , on which the OPTA depends. More precisely, $C = 1 \Leftrightarrow H_2(p)$, where H_2 is the *binary entropy function*.

The Gaussian upper bound is expressed as follows and drawn in figure 2, where we considered $\sigma_u^2 = \frac{1}{12}$ as in a $[\Leftrightarrow \frac{1}{2} \ \frac{1}{2}]$ distributed uniform source:

$$R(D) \leq R^g(D) = \frac{\frac{1}{2} \log_2 \frac{\sigma_u^2}{D}}{C} \quad (2)$$

The $R(D)$ curve of any source is below this bound.

Bitwise decomposition A memoryless source with uniform *probability density function* (pdf) is considered in this section. We

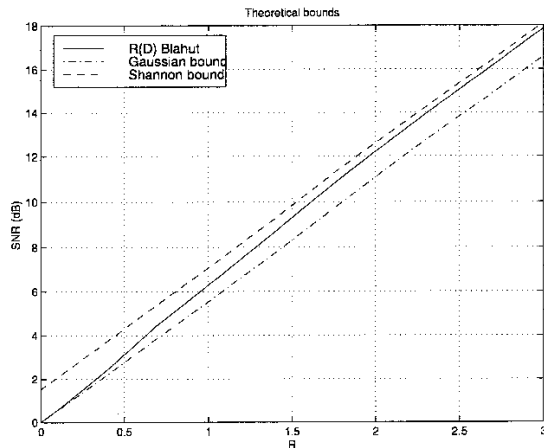


Figure 2: Theoretical bound of rate-distortion curve obtained from the Gaussian source $R^g(D)$; Theoretical Shannon's lower bound of distortion, $R^s(D)$ and $R(D)$ for a uniform source obtained by Blahut's algorithm, all for $p = 0.01$.

see that since different bits have different contributions to the total error, it is rather reasonable to send different bits with unequal compression and/or protection rates. So, the blocks of n_1 *most significant bits* (msb) are grouped together; the blocks of n_2 bits from the next row, until n_N *least significant bits* (lsb) and each bit stream is a BSS [8].

It can be shown that the distortion introduced by each row of bits is additive [8]:

$$D = \sum_{i=1}^N 4^{-i} \mathbf{E}(\mathbf{u}_i \Leftrightarrow \mathbf{v}_i)^2 = \sum_{i=1}^N w_i D_i \quad (3)$$

Lagrangian bound of distortion for a uniform source In [8] we have obtained optimal performance $R_i(D_i)$ for each bit stream i :

$$R_i(D_i) = \frac{1 \Leftrightarrow H_2\left(\frac{1-\sqrt{1-4D_i}}{2}\right)}{1 \Leftrightarrow H_2(p)}. \quad (4)$$

it remains to determine the optimal allocation of bit rates R_i that minimizes the total distortion D , as developed in 3, for a given rate budget $R = \sum_i R_i$. This gives the best performance obtainable for our structure of figure 3. We solve this problem by the Lagrangian multiplier method.

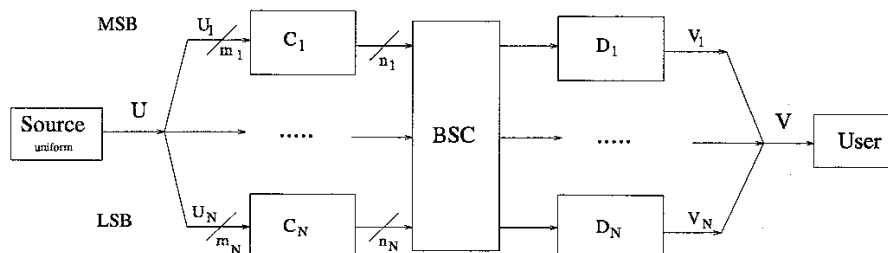


Figure 3: Bitwise decomposition structure: each row of bits (msb, ..., lsb) is coded separately.

The problem is to minimize $R = \sum_{i=1}^N R_i(D_i)$ subject to $D = \sum_{i=1}^N w_i D_i$. This problem has been addressed in [8] and the result is:

$$\frac{1}{\sqrt{1 \leftrightarrow 4D_i}} \log_2 \left(\frac{1 + \sqrt{1 \leftrightarrow 4D_i}}{1 \leftrightarrow \sqrt{1 \leftrightarrow 4D_i}} \right) = \lambda w_i (1 \leftrightarrow H_2(p)) \quad (5)$$

With any positive value of λ , this condition gives the optimal values of the D_i 's. The D_i 's were computed from λ by inverting the complicated function (5) numerically. The result is, for any positive λ , a bit rate $R = \sum_i R_i$ and a value of total distortion $D = \sum_i w_i D_i$ which gives a solution to the problem. This result is drawn in figure 5. It is clearly seen that the best attainable performance of our structure is close to the OPTA curve.

3 Channel Optimized Coder Design for binary sources

Here, we study the case of binary source. The practical algorithms used to design coders for binary input, real-valued outputs are explained. These coders are desirable in our approach since we need to minimize the Euclidean distance of output and input (and not the error probability, as is usually the case).

In our proposition, we aim at minimizing the distortion measure, D , which includes channel noise effect, directly by derivating the expression of D , once with respect to the encoder parameters and once with respect to the decoder parameters. The approach is quite similar to that of Farvardin and Vaishampayan [4] where they propose *Channel Optimized Vector Quantization* (COVQ), as a generalization

of the LBG algorithm for noisy channels. We summarize below, the main features of the algorithm and the results that we use in our further simulations.

We aim at minimizing the average distortion between u and v , with respect to MSE criterion as defined in the following equation:

$$D = \frac{1}{m} \mathbf{E}\{d(\mathbf{u}, \mathbf{v})\} = \frac{1}{m} \mathbf{E}\{\|\mathbf{u} \leftrightarrow \mathbf{v}\|^2\} \quad (6)$$

Equation (7) is obtained for a BSC and is known as *generalized distortion measure* [1]:

$$D = \frac{1}{m} \sum_{\mathbf{u}=0}^{2^m-1} \mathbf{E}_{\mathbf{y}} \|\mathbf{u} \leftrightarrow \mathbf{v}_{\mathbf{y}}\|^2 p(\mathbf{u}) \quad (7)$$

$$= \frac{1}{m} \sum_{\mathbf{x}=0}^{2^n-1} \sum_{\mathbf{u} \in C_{\mathbf{x}}} \sum_{\mathbf{y}=0}^{2^n-1} \|\mathbf{u} \leftrightarrow \mathbf{v}_{\mathbf{y}}\|^2 p(\mathbf{y}|\mathbf{x}) p(\mathbf{u}) \quad (8)$$

where \mathbf{x} is the encoded word to be transmitted on the channel and $C_{\mathbf{x}}$ is the region encoded to \mathbf{x} . In a BSC, the factor $p(\mathbf{y}|\mathbf{x}) = p^{d_H(\mathbf{y}, \mathbf{x})} (1 \leftrightarrow p)^{n-d_H(\mathbf{y}, \mathbf{x})}$ is the conditional probability of observed \mathbf{y} , channel output, given \mathbf{x} as channel input; where d_H denotes Hamming distance. The dependency of distortion to index assignment is due to the $d_H(\mathbf{y}, \mathbf{x})$ factor.

Analogously to LBG, COCD is also an iterative algorithm which makes use of the two following operations: *centroid update* and *boundary update*. These two steps are carried out iteratively until a stability in the distortion is observed.

Centroid update Supposing that the encoder is fixed (the boundaries of the regions), the decoder is updated. As we consider the

channel effect in these expressions, the resulting expression for centroids is called *generalized centroid*. Derivating the expression for distortion, D in respect to \mathbf{v}_y , we have:

$$\sum_{\mathbf{x}=0}^{2^n-1} \sum_{\mathbf{u} \in C_{\mathbf{x}}} 2(\mathbf{u} \Leftrightarrow \mathbf{v}_y) p(\mathbf{y}|\mathbf{x}) p(\mathbf{u}) = 0 \quad (9)$$

$$\mathbf{v}_y = \frac{\sum_{\mathbf{x}=0}^{2^n-1} \sum_{\mathbf{u} \in C_{\mathbf{x}}} p(\mathbf{y}|\mathbf{x}) \mathbf{u} p(\mathbf{u})}{\sum_{\mathbf{x}=0}^{2^n-1} \sum_{\mathbf{u} \in C_{\mathbf{x}}} p(\mathbf{y}|\mathbf{x}) p(\mathbf{u})} \quad (10)$$

This expression says that all input possibilities, $\mathbf{u} \in C_{\mathbf{x}}$ affect all centroids, \mathbf{v}_y and that this effect is weighted by the $p(\mathbf{y}|\mathbf{x})$ factor.

Boundary update Supposing a given decoder (a set of centroids), the encoder is updated. In other words, the boundaries of the regions are updated in order to minimize the distortion. This condition is obtained directly by derivating the expression for distortion, D , and we call it the principle of *generalized nearest neighbor*.

$$\mathbf{x} = \underset{\mathbf{y}=0}{\text{Argmin}}_{\mathbf{x}} \sum_{\mathbf{y}=0}^{2^n-1} \|\mathbf{u} \Leftrightarrow \mathbf{v}_y\|^2 p(\mathbf{y}|\mathbf{x})$$

Again we observe the effect of channel noise in this expression. Presence of the factor $p(\mathbf{y}|\mathbf{x})$ makes the encoding dependent to the channel. In other words, in a highly noisy channel, the encoded information word does not necessarily represent the nearest centroid to the input word, but its generalized nearest neighbor.

We can summarize the whole algorithm as follows:

I. Initialization:

- Consider an initial state for the encoder: $f(\cdot)_0$.

II. Iterations:

For the k -th iteration:

- Update all centroids (decoder $g(\cdot)_k$), given encoder ($f(\cdot)_{k-1}$).

- Update all boundaries (encoder $f(\cdot)_k$), given decoder ($g(\cdot)_k$).

III. Termination:

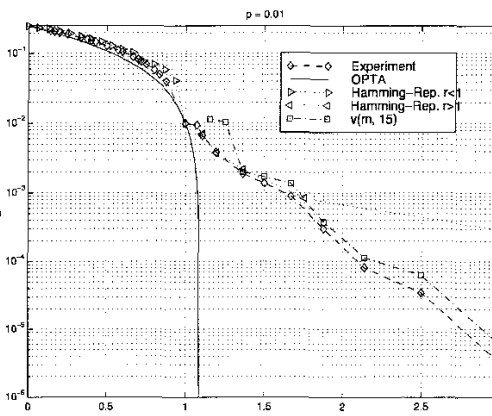


Figure 4: $R(D)$ and the coders used in the practical experiments.

- Calculate distortion D_k ; if $D_k \Leftrightarrow D_{k-1} < \epsilon$, then stop, otherwise repeat (II).

We observe that the sequence of D_k is a decreasing sequence and we can terminate the algorithm when it stabilizes. In our experiments, we observed also that the number of necessary iterations before reaching any local minimum is usually very low (mostly less than five iterations), and that many local minima can be found, according to different initializations.

Practical considerations In practice, there are some problems which limit the application of COCD. The major problem is its complexity. In fact finding a good optimum for the values of m and n greater than, say roughly 15, confronts to practical limitations.

We used the classical coders, as proposed in [2], with $n = 15$, for the initialization of the encoder, $f(\cdot)$, and obtained the local optimum coders. Figure 4, shows the performances of the coders that we have obtained. The figure shows how much these new coders act better than their classical counterparts (Hamming and Repetition codes) and even better than the specifically designed coders proposed in [2] with $n = 15$ ($v(m, n)$).

To complete our set of coders, we added two Hamming coders of slightly longer lengths, $H_{63,57}$, $H_{31,26}$.

The algorithm presented here, works inde-

pendent of the fact that $m < n$ or $m > n$. For $m = n$, the best coding is to do nothing at all! For the $m > n$ case, we observed that the best minima obtained by the algorithm were the *trivial coders*. In a trivial coder, n bits are transmitted and $m \Leftarrow n$ other bits are not transmitted at all. In the receiver, in place of these $m \Leftarrow n$ bits, their mean value (0.5, for a binary source) is introduced. The performance of a trivial coder can be simply obtained as a linear weighting of $D_{(r=1)} = p \Leftarrow p^2$ and $D_{(r=0)} = 0.25$, as follows:

$$D_{\text{trivial}}\left(\frac{n}{m}\right) = \frac{n \times (p \Leftarrow p^2) + (m \Leftarrow n) \times 0.25}{m} \quad (11)$$

In our practical experiments, we consider all trivial coders with the form

$$t(100, n) \quad n = 1, 2, \dots, 99. \quad (12)$$

4 Uniform source

For the uniform source, we considered the same coders obtained for the binary source and explained in last section. For the general structure, we consider the structure in figure 3. We search the optimum R_i, D_i pair for each row, i , in order to minimize the overall distortion, D , as in equation (3), with the constraint that $R = \sum_i R_i \leq R_d$, for a given p .

The optimization was done with the use of the *bit allocation algorithm* as proposed by Shoham and Gersho [6]. This method fits a polyline on the R/D plane. The numerical result is shown in figure 5. For a highly condensed cloud, Shoham's algorithm can deliver the envelope of the permitted region.

Figure 5 shows the result of optimization and compare it to the Lagrangian bound as well as the no-coding bound for $p = 10^{-2}$. The result of a *Channel Optimized Scalar Quantization* (COSQ) are also shown for comparison. It is clear that with COSQ, only integer values of bit rate can be obtained. Also we see that the proposed algorithm overperforms COSQ in this experiment.

Figure 6 illustrates the result of optimization for $p = 10^{-2}$ and for $R_d = 4$ also it compares this result with a system without any

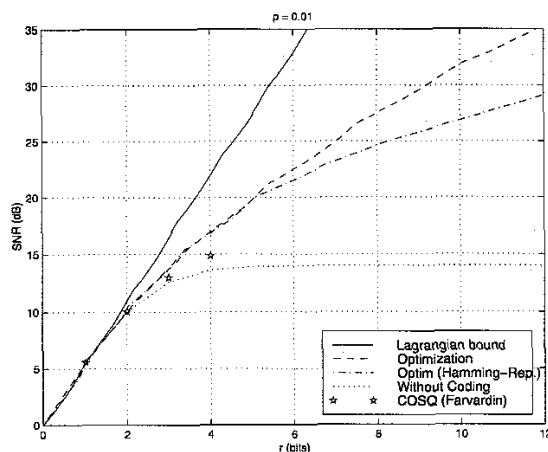


Figure 5: Lagrangian bound, optimization results for the coders used in this simulation (COCD) and those in [7] (Hamming and repetition codes), for $p = 10^{-2}$. Also is shown a curve when no coding is applied and also the results in COSQ.

special coding where all 4 bits are transmitted directly on the channel. One can observe that there is a tendency to equalize the distortion due to each row, $w_i \cdot d_i$, in figure 6-a comparing to figure 6-b. In fact, the space of used coders is quite sparse, otherwise, one could expect a much more equalized error contribution of all lines. Roughly speaking, the distribution of bit rate to each row, R_i , is inversely proportional to the line number i . So: $R_i \geq R_j \iff i \leq j$.

5 Conclusion

Using our new method, COCD, we had significantly improved the performance of a bitwise decomposition system.

It must be indicated however that using more powerful codes might lead more effective optimization. In fact, our experiment can be improved, using a more complete set of coders.

Further work will be dedicated to applying this technique on the generalized Gaussian sources, and then to the set of sources encoded by a transform coder.

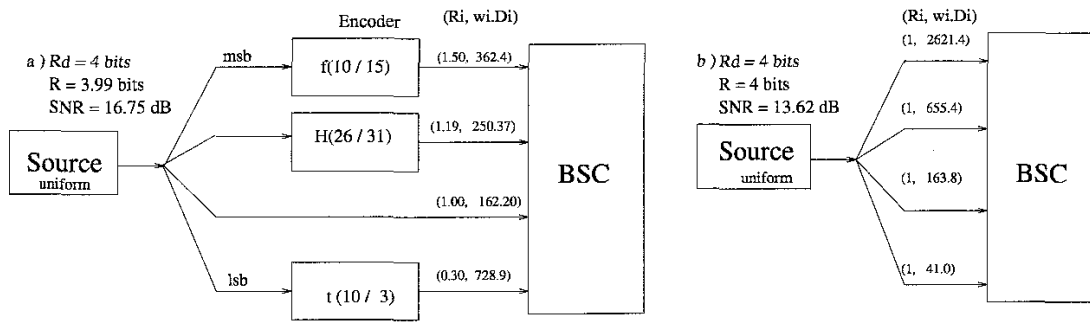


Figure 6: **a)** The optimum system of coders; **b)** A system without coding, for $p = 10^{-2}$ and $R_d = 4$. The pair of numbers in parenthesis are $(R_i, w_i \cdot d_i)$, the bit rate and the contribution to distortion due to each row.

References

1. E. Ayanoglu and R.M. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders", IEEE Transactions on Information Theory **33** (1987), no. 6, 855-865.
2. F.X. Bergot and O. Rioul, "Joint Binary Systematic Source-Channel Coding with Small Linear Codes", Proceedings of the European Conference on Signal Processing and Communications (Rhodes, Greece), Sep. 1998.
3. N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source-Channel Coding", IEEE Transactions on Information Theory **33** (1987), no. 6, 827-838.
4. ———, "On the Performance and Complexity of Channel Optimized Vector Quantizers", IEEE Transactions on Information Theory **37** (1991), no. 1, 155-160.
5. R.J. McEliece, *The Theory of Information and Coding*, Addison-Wesley, 1977.
6. Y. Shoham and A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers", IEEE Transactions on Acoustics, Speech and Signal Processing **36** (1988), no. 9, 1445-1453.
7. S.B. ZahirAzami, P. Duhamel, and O. Rioul, "Combined Source-Channel Coding for Binary Symmetric Channels and Uniform Memoryless Sources", Proceedings of the colloque GRETSI (Grenoble, France), Sep. 1997.
8. S.B. ZahirAzami, O. Rioul, and P. Duhamel, "Performance Bounds for Joint Source-Channel Coding of Uniform Memoryless Sources Using a Binary Decomposition", Proceedings of European Workshop on Emerging Techniques for

Communication Terminals (Toulouse, France), July 1997, pp. 259-263.