Performance Bounds for Joint Source-Channel Coding of Uniform Memoryless Sources Using a Binary Decomposition *

Seyed Bahram ZAHIR AZAMI*, Olivier RIOUL* and Pierre DUHAMEL**

Départements *Communications et **Signal
École Nationale Supérieure des Télécommunications
URA CNRS 820
46, rue Barrault, 75634 Paris cedex 13, France

Abstract

The objective of this paper is to design and evaluate the performance of a transmission system with jointly optimized source and channel coding of a uniformly distributed source to be transmitted over a *binary symmetric channel* (BSC).

We first provide the *optimal performance theoretically attainable* (OPTA) according to Shannon's theorem. Then, we propose a new structure for joint source and channel coding in which the samples are first expressed in binary representation and bits of same weight are processed separately. Finally, we determine the lower bound for total distortion attainable with this structure and compare it to OPTA curves and to simulation results.

1 Introduction

This paper addresses the transmission of digital data over noisy channels with jointly optimized source and channel coders. The type of source considered in this paper has a *uniform* probability distribution, and the transmission is over a *binary symmetric channel* (BSC). Our choices for the source and channel may seem overly simplistic. We feel, however, that their study gives a better understanding of the problem of joint source/channel coding. Moreover, we shall derive building blocks that can be used in more sophisticated systems.

Figure 1 illustrates our transmission system. In this configuration, the source/channel code rate is defined as the average number of coded bits per source sample: $r = \frac{n}{m}$. We seek to minimize the m.s.e. distortion $D = \frac{1}{m} E\{\|U - U'\|^2\}$, with the constraint that $r \le r_d$, where r_d is the desired bit rate.

*This work is supported by a grant from the CNET-France Télécom. The work of S.B. Zahir Azami is also sponsored by the CNOUS.

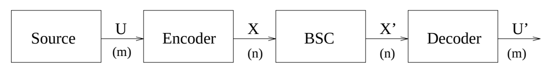

Figure 1: A simple transmission structure.

From Shannon theory we know that source and channel coding can be treated separately without any loss of performance for the overall system [3]. This is, however, an asymptotic result, as it necessitates very long blocks and very complex coders. Our approach is to achieve relatively good results, using comparatively simple coders.

In this paper, we first derive upper and lower bounds for the optimal rate-distortion functions, taking channel transmission errors into account. Then, we propose a new structure in which the source is decomposed into parallel binary streams. For this structure, we find a theoretically attainable rate-distortion function, using the Lagrangian multiplier method, and compare it to the optimal one. Finally, we compare these theoretical results with those obtained by a practical method based on the proposed structure.

2 Bounds for Uniform source

A memoryless source with uniform probability density function is considered. This source is to be coded and transmitted over a BSC. Note that the uniformity of the source does not permit too much for source coding, except for the dimensionality that can be exploited [2]. We now proceed to derive the optimal rate-distortion function $r(D)$, as well as upper and lower bounds.

2.1 OPTA

The *optimal performance theoretically attainable* (OPTA) 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 [3], the OPTA curve $r(D)$ is given by

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

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

For our model, the BSC is parameterized by the *raw bit-error probability* p, on which the OPTA depends. More precisely, one has $C = 1 - H_2(p)$, where $H_2(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$ is the binary entropy function. There is no closed-form expression for the uniform source $R(D)$ function; it is derived below with the aid of Blahut's algorithm [1]. However, it is a simple matter to obtain closed-form expressions for lower and upper bounds, as shown next.

2.2 Gaussian upper bound

It is known that a theoretical upper bound of the source rate-distortion function $R(D)$ (without channel consideration) is given by the source rate-distortion function $R_g(D)$ of a Gaussian source of same variance σ^2 [5]:

$$R(D) \leq R_g(D) = \frac{1}{2} \log_2 \frac{\sigma^2}{D}$$

This gives an upper bound for the OPTA curve:

$$r(D) \leq r_g(D) = \frac{\frac{1}{2} \log_2 \frac{\sigma^2}{D}}{1 - H_2(p)} \quad (2)$$

This upper bound is plotted in figure 2 for a uniform source distributed on $[-\frac{1}{2}; \frac{1}{2}]$, for which $\sigma^2 = \frac{1}{12}$.

2.3 Shannon's lower bound

The Shannon's lower bound [5] for the source rate-distortion function is given by

$$R(D) \geq R_s(D) = H - \frac{1}{2} \log_2 2\pi e D.$$

where $e = 2.71828...$ and H denotes the differential entropy of the source, given by $H = \frac{1}{2} \log_2 12\sigma^2$ for a uniform source. It has been observed that this bound is not attainable except for very small distortions.

Shannon's theorem gives a lower bound for the OPTA curve:

$$r(D) \geq r_s(D) = \frac{\frac{1}{2} \log_2 \frac{12\sigma^2}{2\pi e D}}{1 - H_2(p)} \quad (3)$$

This lower bound is plotted in figure 2.

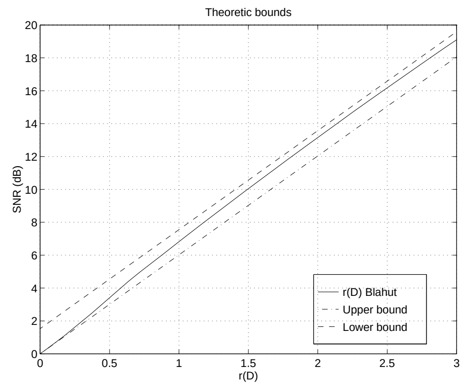

Figure 2: Upper and lower theoretical bounds, compared with the OPTA curve obtained by Blahut's algorithm, for a uniform source and zero error channel ($p = 0$).

2.4 Blahut's algorithm

Using Blahut's algorithm [1], we found numerically the source-distortion function $R(D)$. From (1) this gives the OPTA curve $r(D) = \frac{R(D)}{1-H_2(p)}$ plotted in figure 2. We observe that the OPTA curve is close to the upper bound for the small values of r and comes closer to the lower bound for the large values of r. Note that the distance between the lower and upper bounds is constant, about 1.5 dB.

3 A new joint source-channel coding structure

3.1 Bitwise decomposition

The bounds derived above are useful since one may check the performances of practical algorithms against them. On the other hand, we are searching for simple algorithms and have chosen to decompose a uniform source U into a set of N binary sources U_i which are transmitted through the same channel, as shown in figure 3. Thus, we consider the binary representation of each source sample, truncated on N bits, and then perform the compression and the protection operations on each bit stream separately.

An important consideration is the following. We assume that the original source U is uniformly distributed (e.g., between $-1/2$ and $1/2$) and that natural binary coding is used, i.e., the U_i's are such that $U = \sum_{i=1}^{N} 2^{-i} U_i$. Then it is easily seen that each bit stream U_i is a *binary symmetric source* (BSS), that is, the bits U_i are independent and identically distributed with $\text{Prob}(U_i = 0) = \text{Prob}(U_i = 1) = 1/2$.

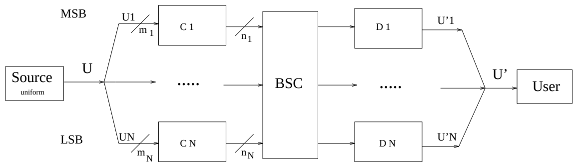

Figure 3: Source-Channel coder combination. Each row of bits (msb, ..., lsb) is processed separately.

Of course, the output will be reconstructed by the formula $U' = \sum_{i=1}^{N} 2^{-i} U'_i$.

3.2 Remark on memoryless sources

Before deriving optimal performances, a preliminary remark is in order. It follows from information theory that for any memoryless source, the rate-distortion function may be determined as $R(D) = \min_{p(u)}\{I(U,U'); \mathbf{E}(U-U')^2 \leq D\}$ where U and U' represent input and output random variables (not vectors) and $I(U,U')$ is the mutual information between U and U'. Therefore, even though actual processing is made by blocks of length m, the calculation of $R(D)$ is made using the definition of D for $m=1$, that is, $D = \mathbf{E}(U-U')^2$.

3.3 Additivity of distortion

A second important consideration is that distortion is additive. To prove this, write

$$\begin{aligned} D &= \mathbf{E}\,(U-U')^2 \\ &= \mathbf{E}\left(\sum_{i=1}^{N} 2^{-i} U_i - \sum_{i=1}^{N} 2^{-i} U'_i\right)^2 \\ &= \sum_{i,j} 2^{-(i+j)} \mathbf{E}\,(U_i - U'_i)(U_j - U'_j) \end{aligned}$$

Since U_i and U_j are independent for $i \neq j$, and considering that U_i and U'_i have same biases (that is, $\mathbf{E}\,(U_i) = \mathbf{E}\,(U'_i)$) we have $\mathbf{E}\,(U_i - U'_i)(U_j - U'_j) = \mathbf{E}(U_i - U'_i)\mathbf{E}(U_j - U'_j) = 0$ for all $i \neq j$. Therefore, D simplifies to

$$D = \sum_{i=1}^{N} 4^{-i} \mathbf{E}\,(U_i - U'_i)^2 = \sum_{i=1}^{N} w_i D_i$$

where $D_i = \mathbf{E}\,(U_i - U'_i)^2$ is the m.s.e. corresponding to bit i, which depends on the channel raw error probability p, and $w_i = 4^{-i}$ is the weighting associated with bit i.

Thus, D is a linear superposition of bit distortions D_i. It is important to note that this result was obtained with the assumption that $\mathbf{E}\,(U_i) = \mathbf{E}\,(U'_i)$. Since we have decomposed our source into BSS sources, the next step is to find the OPTA performance $r_i(D_i)$ for each BSS U_i transmitted over a BSC. This is done in the following section.

4 OPTA for a binary source

4.1 Error probability distortion

The OPTA for a BSS U_i over a BSC is well known when the distortion δ_i is defined as an error probability: $\delta_i = \mathbf{E}\,(w_H(U_i - U'_i))$ where $w_H(x)$ is the Hamming weight of x, that is, $w_H(x) = 0$ if $x = 0$ and $= 1$ otherwise. This clearly corresponds to an m.s.e. distortion $\mathbf{E}(U_i - U'_i)^2$ if we require that U_i and $U'_i \in \{0,1\}$. In this case the source rate-distortion function is given by [3, 5]

$$R_i(\delta_i) = 1 - H_2(\delta_i)$$

From (1) this gives the following OPTA curve.

$$r_i(\delta_i) = \frac{1 - H_2(\delta_i)}{1 - H_2(p)} \tag{4}$$

4.2 m.s.e. distortion

Even though our basic system is binary, we want the additional flexibility in the output U'_i to be different from $\{0,1\}$, because it yields lower values of m.s.e. distortion and also permits the requirement that $\mathbf{E}(U_i) = \mathbf{E}(U'_i)$, which was needed to obtain the additivity property in the preceding section.

Consider, for example, the limit case $r_i = 0$. Then $U'_i = 1/2$ is the value that minimizes $D_i = \mathbf{E}(U_i - U'_i)^2$, and one has $\mathbf{E}(U_i) = \mathbf{E}(U'_i)$. In general, the output is still reconstructed as $U' = \sum_{i=1}^{N} 2^{-i} U'_i$ even though the U'_i are no longer "bits".

With this assumption, we have to re-calculate the OPTA for a m.s.e. measure of distortion D_i in place of the error probability distortion δ_i. But there is a simple relationship between them, which we now derive.

Replace $U'_i = 0$ by c_1's and the $U'_i = 1$ by c_2 in the output, where c_1 and c_2 are real-valued constants. We seek to minimize D_i for a given δ_i with a good choice of these constants. Since

$$D_i = (1-\delta_i)\frac{1}{2}\{c_1^2 + (1-c_2)^2\} + \delta_i \frac{1}{2}\{(1-c_1)^2 + c_2^2\},$$

it is easily seen that $c_1 = 1 - c_2 = \delta_i$ is the choice that minimizes D_i, and we have our basic requirement that $\mathbf{E}(U_i) = \mathbf{E}(U'_i) = 1/2$. It follows that

$$D_i = \delta_i - \delta_i^2,$$

which determines the value of the error probability $\delta_i \leq \frac{1}{2}$ that should be used in the OPTA (4) as

$$\delta_i = \frac{1}{2}(1 - \sqrt{1 - 4D_i}).$$

The OPTA curve is, therefore, given by:

$$r_i(D_i) = \frac{1 - H_2(\frac{1-\sqrt{1-4D_i}}{2})}{1 - H_2(p)}. \quad (5)$$

In figure 4, we have plotted the corresponding source rate-distortion function $R_i(D_i) = 1 - H_2(\frac{1-\sqrt{1-4D_i}}{2})$, along with operating points obtained for simple coders used in the simulation presented in [6].

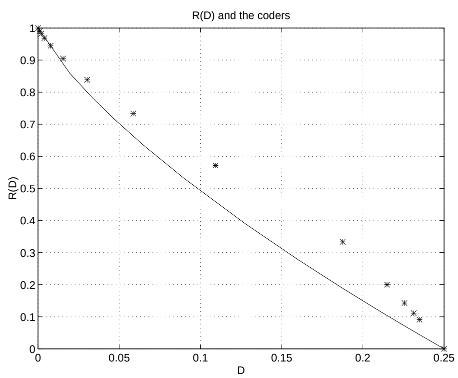

Figure 4: $R_i(D_i)$ and simple repetition and Hamming coders used in our simulations.

Figure 5 illustrates the behavior of the OPTA distortion D_i as a function of the BSC raw error probability p, for different values of bit rate r_i. Several remarks are in order.

- For $r_i = 0$ (no transmission at all) we have an horizontal line $D_i = \frac{1}{4}$ which corresponds to $U_i' = \frac{1}{2}$ at the receiver, as we have already noticed.
- For $r_i = 1$, we have $D_i = p - p^2$ and $\delta_i = p$. The total error probability distortion is imposed by the channel, and the optimal choice is to do no coding at all.
- For $r_i = \infty$, we obtain the vertical line $p = \frac{1}{2}$. This corresponds to the limit case $\lim_{r_i \to \infty} D_i = 0$ for all p.

5 Lagrangian bound

Now that we have determined the optimal performance $r_i(D_i)$ for each bit stream i, it remains to determine the optimal allocation of bit rates r_i that minimizes the total distortion $D = \sum_i w_i D_i$ for a given rate budget

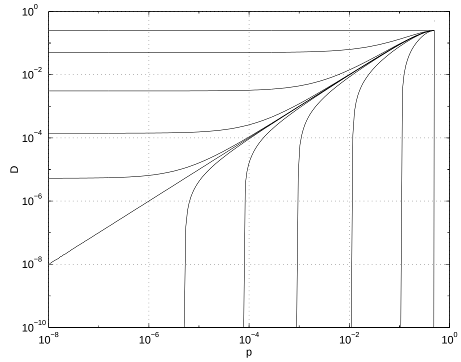

Figure 5: Theoretical bound for distortion function versus row-error probability, p, for different values of $r_i \in \{0, 0.7, ..., 0.9999, 1, 1.0001, ..., 1.1, 2, \infty\}$.

$r = \sum_i r_i$. This will give the OPTA $r(D)$ for our structure of figure 3. We solve this problem by the Lagrangian multiplier method.

5.1 Derivation

The problem is to minimize $D = \sum_{i=1}^{N} w_i D_i$ subject to $\sum_{i=1}^{N} r_i = r$ where $r_i(D_i)$ is given by (5). An equivalent problem in the variables $D_1, \ldots, D_N$ is to minimize $r = \sum_{i=1}^{N} r_i(D_i)$ subject to $\sum_{i=1}^{N} w_i D_i = D$. For convenience in the derivation we use the latter form of the problem.

Let λ be the Lagrange multiplier corresponding to the constraint. Since the objective function r to be minimized is convex, and the constraint is linear in the D_i's, the solution to our problem is determined by the equation

$$\frac{\partial \mathcal{L}}{\partial D_i} = 0$$

where $\mathcal{L}$ is the Lagrangian functional

$$\mathcal{L} = \sum_i r_i(D_i) + \lambda(\sum_i w_i D_i - D)$$

Using (5) and the formula $\frac{\partial H_2(x)}{\partial x} = \log_2(\frac{1-x}{x})$ we obtain the following necessary and sufficient condition for optimality:

$$\frac{1}{\sqrt{1-4D_i}} \log_2(\frac{1+\sqrt{1-4D_i}}{1-\sqrt{1-4D_i}}) = \lambda w_i(1 - H_2(p)) \quad (6)$$

Now, with any value of $\lambda > 0$, this condition gives the optimal values of the D_i's. These in turn give the values of the r_i's from (5). The D_i's were computed from λ by inverting the complicated function (6) numerically. The result is, for any $\lambda > 0$, a bit rate $r = \sum_i r_i$ and a value of total distortion $D = \sum_i w_i D_i$ which give a solution to the problem.

5.2 Results and comparison

The result (which we call "Lagrangian bound") is plotted in figure 6. Notice that the curves obtained for different values of p are just scaled versions of each other on the r axis.

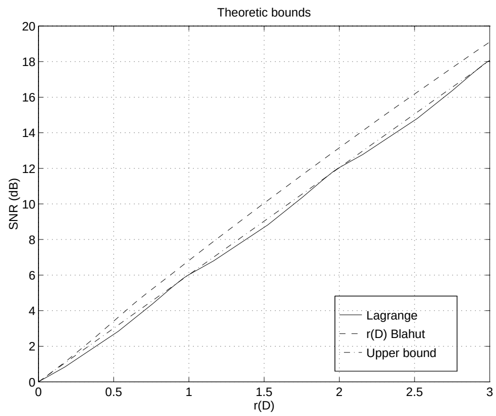

Figure 6: **Solid**: Theoretical OPTA curves (Lagrange, Blahut) and the upper bound for $p = 0$.

Figure 6 also compares our Lagrangian bound to the OPTA curve obtained by Blahut's algorithm and the Gaussian upper bound (2). It is interesting to note that the Lagrangian bound is tangent to the upper bound every time R is an integer. This can be explained as follows.

It turns out that the upper bound (2) is also, when R is an integer, the optimal rate-distortion function when *scalar quantization* is used on the source samples U. Indeed, assuming for example that the source is uniformly distributed on the interval $[-\frac{1}{2}; \frac{1}{2}]$, so that $\sigma^2 = \frac{1}{12}$, it is easy to see that $R = R_g(D)$ in (2) is equivalent to $D = \sigma^2 2^{-2R} = \frac{q^2}{12}$ where $q = 2^{-R}$ is the quantization step. Now, by using our bitwise decomposition as in figure 3, even though bits are processed block-wise, it seems that we have lost the ability to *vector quantize* the source samples U. Note, however, that our structure was proposed for a joint source/channel coding problem for which one also considers transmission errors due to a *binary* channel. It would be desirable to improve our structure to permit vector quantization while also taking binary transmission errors into account. This is a subject for future investigation.

In [6], we have proposed an optimization procedure based on the proposed bitwise structure, using simple binary coders. The rate-distortion optimization was performed using a variation of Shoham and Gersho's algorithm [4]. Figure 7 shows the result of this optimization and compares it to the Lagrangian bound.

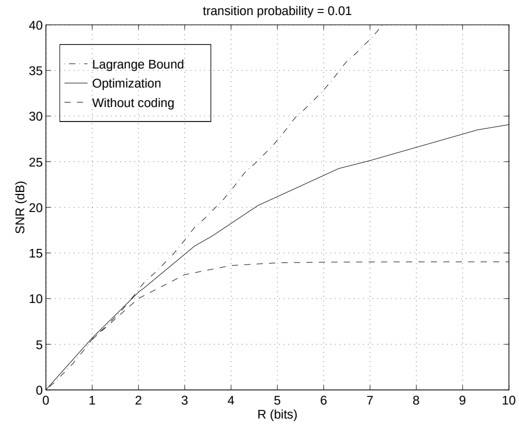

Figure 7: Numerically attainable bound (with the used set of coders) and theoretical bound obtained by Lagrange's multiplier method. Also shown is the performance curve obtained without any coding.

6 Conclusion

In this paper we have proposed a new structure for joint source and channel coding of a uniform source to be transmitted over a binary symmetric channel, and have derived precise theoretical performance for this structure.

Further possible improvements include the ability to vector quantize the source samples on their binary representation, to take other types of sources into account (including Gaussian and Laplacian) and to treat the case of multiple sources.

References

1. R.E. Blahut, *"Computation of chanel capacity and rate-distortion functions"*, IEEE Transactions on Information Theory **18** (1972), no. 4, 460–473.
2. J. Makhoul, S. Roucos, and H. Gish, *"Vector quantization in speech coding"*, Proceedings of the IEEE **73** (1985), no. 11, 1551–1588.
3. R.J. McEliece, *The theory of information and coding*, Addison-Wesley, 1977.
4. 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.
5. A.J. Viterbi and J.K. Omura, *Principles of digital communication and coding*, McGraw-Hill, second ed., 1985.
6. 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, to appear.