Combined Source-Channel Coding for Binary Symmetric Channels and Uniform Memoryless Sources

S.B. Zahir Azami, P. Duhamel and O. Rioul

École Nationale Supérieure des Télécommunications
46, rue Barrault, 75634 Paris Cedex 13, France
Tél: +33 1 45 81 76 77, Fax: +33 1 45 89 00 20, email: zahir@com.enst.fr

RÉSUMÉ

On étudie une conception conjointe de codage source et canal pour la transmission numérique de données sur un canal bruité. Nous proposons un algorithme débit/distortion pour obtenir la meilleure combinaison possible de codeurs source et canal parmi un ensemble donné, afin de minimiser la distorsion totale pour un débit source/canal donné. La complexité du système global reste raisonnable grâce à l'utilisation de briques de base simples. On utilise une technique similaire à celle de Shoham et Gersho, ce qui est rendu possible par l'utilisation d'un schéma original, dans lequel chaque train de bits (msb, ..., lsb) est codé séparément. On compare les performances obtenues aux performances optimales prévues dans les mêmes conditions par la théorie de Shannon.

ABSTRACT

A joint design of source and channel coding is considered for digital transmission over a noisy channel. We propose a rate/distortion algorithm for obtaining the best possible combination of a given set of source and channel coders, so as to minimize the total distortion for a given source/channel bit rate budget. The overall system complexity is maintained low by using simple building blocks. We use a technique similar to the classical Shoham and Gersho algorithm. This is made possible by exploiting an original coding scheme in which each row of bits (msb, ..., lsb) is coded separately. Our results are compared to the optimal performance theoretically attainable (OPTA) according to Shannon theory.

1 Introduction

This paper addresses the transmission of digital data over noisy channels with jointly optimized source and channel coders. The sources considered in this paper are *binary symmetric source* (BSS) and uniform source. The transmission is considered to be over a *binary symmetric channel* (BSC).

These choices of source and channel may seem overly simplistic, but are studied first for a better understanding of the problem, and will later provide building blocks to be used in a more sophisticated system.

Actual communication systems carry sources of various types, which require different error protections. For that 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 will be transmitted. The additional protection is source-dependent, and can be merged with the source coder. By doing so, one can perform joint source/channel coding without dedicating the system to some specific source.

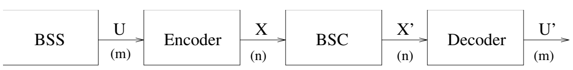

FIG. 1 — A BSS, coded and transmitted over a BSC.

A digital transmission system is presented in Figure 1. In this configuration, the source/channel 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 the quadratic distortion $D = E\{\|U - U'\|^2\}$, with the constraint that $r \leqslant r_d$, where r_d is the *desired bit rate*.

It has long been usual to do source coding and channel coding separately, owing to Shannon's source-channel coding theorem. This theorem consists of two parts [4, 5] : The direct part states that if the minimum achievable source coding rate of a given source is below the capacity of a channel, then the source can be reliably transmitted through the channel. This requires the use of appropriately long blocks of source samples. The converse part states that if the source coding rate is greater than the channel capacity, then a reliable transmission is impossible.

As a result of Shannon theorem, source coding and channel coding can be treated separately without any loss of performance for the overall system. This is an asymptotic result, as it necessitates very long blocks and very complex coders. Our approach is to achieve some relatively good results, using comparatively simple coders.

2 Theoretical bound of distortion for a binary source

In this section we investigate the case of a *binary symmetric source* (BSS) ;

First recall basic results concerning the source-distortion function, $R(D)$, which gives a lower bound on source rate R_s for a given distortion D, without considering any transmission (or, equivalently, noiseless transmission)[4] :

$$R_s \geqslant R(D) = 1 - H_2(D) \quad (1)$$

where $H_2(D)$ is the *binary entropy function* [2, 3].

Also Shannon's channel coding theorem states that with a channel coder rate R_c less than channel capacity C, and for a BSC raw error probability p, a reliable transmission is possible. This means [4] :

$$R_c \leqslant C = 1 - H_2(p) \quad (2)$$

Combining equations (1) and (2) we obtain the *optimum performance theoretically attainable* (OPTA) on a *binary symmetric channel* (BSC), given by :

$$\begin{aligned} r = \frac{n}{m} = \frac{R_s}{R_c} &\geqslant \frac{1 - H_2(D)}{1 - H_2(p)} \\ r - r.H_2(p) &\geqslant 1 - H_2(D) \\ H_2(D) &\geqslant 1 - r + r.H_2(p) \\ D &\geqslant D(r) = H_2^{-1}(1 - r + r.H_2(p)) \quad (3) \end{aligned}$$

This bound $D(r)$ serves as a reference for our further work.

3 R/D allocation - binary source

The scheme we consider has been selected in order to allow the use of the same building blocks for source coding (SC) as well as channel coding (CC) : Using the duality that exists between source coding and channel decoding (CD), we have considered CD as SC.

The optimization is done by selecting the couple (SC, CC) which produces the minimum distortion while keeping the total rate within the required limit, $R_s/R_c \leq r$. In this study, all coders were selected from a limited set, made from 4 simple subsets defined as : $\mathcal{N} \cup \mathcal{R} \cup \mathcal{H} \cup \mathcal{U}$, where :

- Null code : $\mathcal{N} = \{N_{(1,0,0)}\}$,
- Repetition codes : $\mathcal{R} = \{R_{(m,\,1,\,m)}\}$ $m = 3, 5, ..., 11$,
- Hamming codes : $\mathcal{H} = \{H_{(2^m-1,\,2^m-1-m,\,3)}\}$ $m = 3, 4, ..., 10$,
- Universal code : $\mathcal{U} = \{U_{(1,1,1)}\}$.

As a consequence, the attainable rates (either on the source, or on the channel) using these coders are :

$$\mathcal{A} = \{0, \frac{1}{11}, \frac{1}{9}, \frac{1}{7}, \frac{1}{5}, \frac{1}{3}, \frac{4}{7}, \frac{11}{15}, \frac{26}{31}, \frac{57}{63}, \frac{120}{127}, \frac{247}{255}, \frac{502}{511}, \frac{1013}{1023}, 1\}$$

and, the available source/channel rates are $r = \frac{R_s}{R_c}$, where $R_s, R_c \in \mathcal{A}$. In this case, the search is done exhaustively among all possible SC-CC combinations. The number of searches is $(\text{Card}\{\mathcal{A}\})^2 = 15^2 = 225$. Figure 2 depicts the OPTA curve as well as the obtained curves for some raw error probabilities p. For $r = 1$, we have $D = p$ in all the cases (circles on the Figure 2). This is a quite logical result which is easily checked by simulation.

For $p = 10^{-5}$, the two curves practically coincide. It means that we can reliably transmit the data, using our simple coders. When p increases, the two curves deviate sooner and sooner. For $p = 10^{-4}$, these very simple coders still allow an improvement of the distortion by a factor approaching 100, while for $p = 10^{-3}$ and $p = 10^{-2}$ the deviation begins much sooner. This is a consequence of our choice of very simple coders, since the Hamming codes cannot correct more than one bit error in a block. In fact, for high raw error probabilities, the probability of having more than one bit error in a block becomes important, and the Hamming code becomes inefficient. The other codes (repetition codes) can perform better but rapidly increase the rate ($r \geq 3$). The effect of the repetition codes is clearly seen on Figure 2-d ($p = 10^{-2}$) as the steps on the distortion curve for $r = 3, 5, 7, 9$. Obtaining better curves requires the use of more powerful codes, which should be included in $\mathcal{A}$.

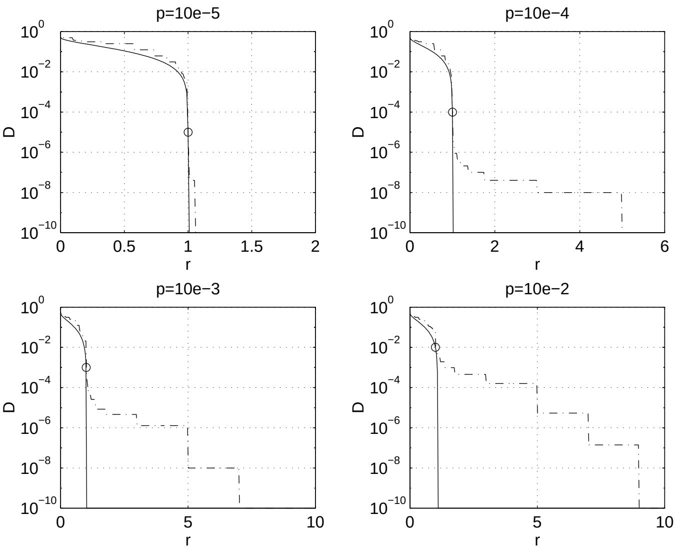

FIG. 2 — **Solid :** OPTA curve D(r). **Dashed :** Points obtained in simulation using Hamming and Repetition codes. For $p = 10^{-5}$ (a), 10^{-4} (b), 10^{-3} (c), 10^{-2} (d).

4 R/D allocation - uniform source

Generalizing the previous section, a memoryless source with uniform pdf is considered. This source is coded and transmitted over a BSC. For such uniform and white source, only dimensionality can be exploited for compression [1]. However, we will see that since different bits have different contributions to the total error, some tradeoff can be obtained : the bits with small influence on the distortion can be transmitted with many errors (rough source quantization), in order to save some bit rate which can be allocated for better protection of more significant bits (efficient error correction).

This requires a separate treatment for bits of different weights. Hence, we consider the successive samples as a set of successive bits, on which the procedure described in the previous section for BSS is applied. This is depicted on fig. 3. Notice that our bit stream is somewhat unusual, since we constitute blocks of bits of same weight : n_1 bits among the *most significant bits* (msb), n_2 bits from the next row, ..., n_N among *least significant bits* (lsb).

This simple architecture allows a different combination of source and channel coders for each row, according to the respective influence of that row on the overall distortion.

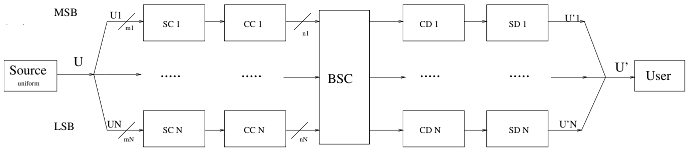

FIG. 3 — Source-Channel coder combination. Each row of bits (msb, ..., lsb) is treated separately.

Considering Card$\{\mathscr{A}\} = 15$, the total number of possible combinations grows exponentially with the number of bits used. Fixing the maximum number of bits, $N = 10$, an exhaustive search would choose the best combination among $15^{10\times2} = 3.32 \times 10^{23}$ possibilities. This is clearly an impossible task.

However, the distortion can be shown to be *additive* (with appropriate weights) on the distortions due to each line. In other words :

$$D = \sum_{i=1}^{N} 4^{N-i} D_i \quad (4)$$

where D_i is the distortion introduced by the bit of weight i. Now, since the distortion, as well as the bit rates, are additive, one can use a procedure similar to the *bit allocation algorithm* of Shoham and Gersho (SGA) [6]. This method searches for the convex envelope of all possible coder combinations in the R/D plane. the method is known to be optimal, provided that the density of attainable points is sufficiently high along the envelope. The additivity allows the multidimensional search to be performed as a succession of mono-dimensional searches. Thus, the complexity is drastically reduced.

Some numerical results are shown on Figure 4, which provides, for a raw error probability $p = 0.01$:

– the OPTA curve

– the envelope of all possible combinations

– a cloud depicting some random SC/CC combinations.

The cloud is clearly a small proportion of all 3.32×10^{23} possibilities. This emphasizes the very small density of available coders in the vicinity of the envelope. The comparison of the envelope curve with the theoretical attainable bound shows that for small word lengths, and small bit rates ($r \leqslant 5$), the attainable R/D curve closely follows the optimal one, despite the simplicity of our coders.

The SGA algorithm iteratively uses a subroutine working on a bit by bit basis : Given the slope λ of a line, it estimates the point of tangency of the line with the cloud of possible combinations. then, SGA searches for λ_{opt} which yields a bit rate less than (or equal to) R_d, while providing minimum distortion. The algorithm iteration is similar to a binary search ; see Figure 5.

The complete algorithm is summarized as follows :

1. Two bounds of tangent are guessed : superior bound, λ_1 and inferior bound, λ_2. It is supposed that the searched point has a tangent, λ, which satisfies : $\lambda_2 \geq \lambda \geq \lambda_1$.
2. Main iteration to find (R, D).

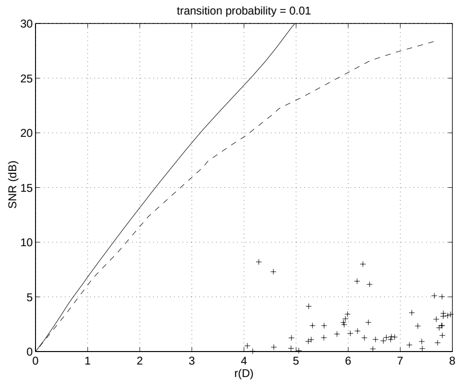

FIG. 4 — The cloud of possible combinations. **Solid** : OPTA curve ; **Dashed** : the performance of R/D coder.

   (a) Two points on the envelope are obtained : (r_1, d_1), (r_2, d_2) corresponding respectively to the tangents λ_1, λ_2.
   (b) The new value for λ^* is guessed as $\lambda = \frac{d_1 - d_2}{r_1 - r_2}$.
   (c) For this new λ^*, the corresponding tangent point (R^*, D^*) is found.
   (d) If $R^* > R_d$ then λ^* is replaced in λ_1 otherwise it is replaced in λ_2.
   (e) The iterations are repeated till there is no more change in the R^* value.
3. A final fine-tuning search is also applied to find the best solution.

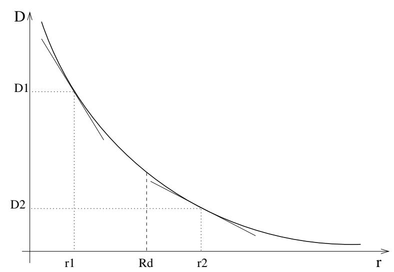

FIG. 5 — The binary search for optimal tangent to the curve consists of defining $\lambda^* = \frac{d_1 - d_2}{r_1 - r_2}$.

The fine tuning step refers to solve the problem introduced by the small density of available points in the vicinity of the best combinations. In this case, it is useful to find the "hidden points", i.e., the points that are not on the envelope but are sometimes better than those on it. For instance, Figure 6 shows for $p = 10^{-2}$ a hidden point (3.95, 19.48) which is between the two points on the envelope, (4.05, 19.79) and (3.88, 19.25). Evidently from the figure, for $R_d = 4$, SGA could yield the (3.88, 19.25) point which is not as good as the hidden point in this example. This point is found in the fine-tuning search and is in fact the global optimal point, too (according to the exhaustive search done in this example). The hidden points are obtained by exhaustive search in the vicinity of the current SGA result.

As explained, the SGA is not guaranteed to give the global optimum for sparse density codes, however we always

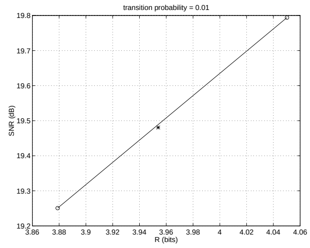

FIG. 6 — A hidden point (⋆) between two points on the envelope (○).

obtained very good results for almost all cases. For example, a simplified exhaustive search resulted in an SNR of 19.48 dB, for $p = 10^{-2}$ and $R_d = 4$ bits, which is the same result obtained by the SGA.

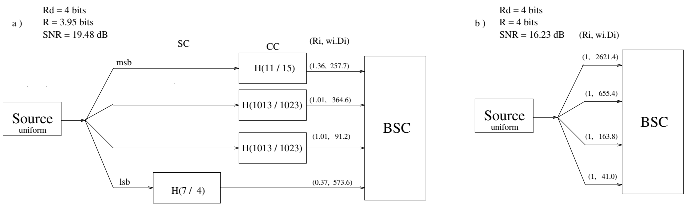

FIG. 7 — **a)** The optimum system of coders ; **b)** A system without SC or CC coding, for $p = 10^{-2}$ and $R_d = 4$. The pair of numbers in parenthesis are $(r_i, w_i \times d_i)$, the bit rate and the contribution to distortion due to each row.

Figure 7 provides the bit rates as well as the contribution of each line of bits to the total distortion for $p = 10^{-2}$ and for $R_d = 4$ in the two following situations : (a) optimum result ; (b) a system without any 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, $w_i \times d_i$, in Figure 7-a comparing to Figure 7-b. The distribution of bit rate to each row, r_i, decreases as the line number i increases, i.e., as weight decreases.

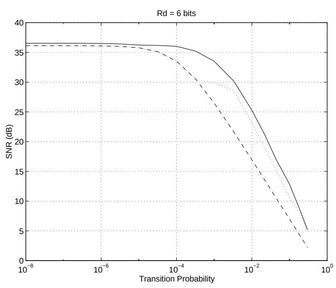

FIG. 8 — The system performances : **Solid** : rate/distortion optimized system ; **Dashed** : without any coding ; **Dotted** : with the same channel coding for all the rows.

The performance of our optimized system is compared to two other system performances in Figure 8 : a system without any coding at all ; and a system with only one CC for all bits. All the systems run at 6 bit/sample rate.

As shown in Figure 8, the two last systems have practically the same performances for transition probabilities below $p < 2 \times 10^{-4}$ while the performances of the proposed optimizations are always at least 1 dB above the two others. The maximum gain is about 8.5 dB. It is also noticeable that even for non noisy channels ($p \to 0$) the proposed algorithm provides some gain. This can be roughly explained in Figure 9-a. As we see, even though $R_d = 3$, there are 4 bits actually in use. This is due to the compression made on the third row, which allows some flexibility to transmit the fourth bit. The overall system performs better than the system on Figure 9-b which contains no coding at all.

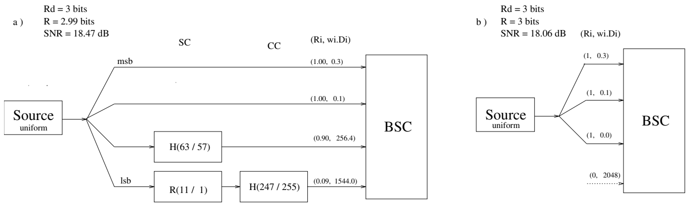

FIG. 9 — a) The optimum system of coders and b) a system without SC or CC coding, for $p = 10^{-6}$ and $R_d = 3$. The pair of numbers in parenthesis are $(r_i, w_i \times d_i)$, the bit rate and the contribution to distortion due to each row.

5 Conclusion

This paper has proposed a rate-distortion optimization algorithm for combined source/channel coding. It is based on a novel coding structure, in which each bit is treated separately.

Although the model used for this system is very simple, the optimization provides useful results in a wide range of bit rates. Further improvement should be obtained using more complex coders.

The most difficult point is that we require distortions to be additive. This is a strong condition on the signal. We are currently working on extensions of the method allowing to take this problem into account. Further work will be reported.

Références

[1] J. Makhoul, S. Roucos and H. Gish, "Vector quantization in speech coding," *Proceedings of the IEEE*, vol. 73, no. 11, pp. 1551–1588, Nov. 1985.

[2] R.J. McEliece, *The theory of information and coding*, Addison-Wesley, 1977.

[3] J.G. Proakis, *Digital communications*, McGraw-Hill, third edition, 1995.

[4] C.E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3&4, pp. 379–423 and 623–656, July and Oct. 1948.

[5] C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 4, pp. 142–163, Mar. 1959.

[6] Y. Shoham and A. Gersho, "efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.