Joint binary symmetric source-channel coding with small linear codes

François-Xavier Bergot, Olivier Rioul

JOINT BINARY SYMMETRIC SOURCE-CHANNEL CODING WITH SMALL LINEAR CODES

François-Xavier Bergot and Olivier Rioul
URA CNRS 820
Communications and Electronics Dept. (COMELEC), E.N.S.T.
46, rue Barrault
75634 Paris CEDEX 13, FRANCE
Emails: bergotfx@com.enst.fr rioul@com.enst.fr

ABSTRACT

The strong similarity between binary symmetric source (BSS) coding relative to Hamming distortion and binary symmetric channel (BSC) coding is discussed in this paper. It is further utilized for designing data compression systems with error probability criterion. These are, in turn, used to construct simple joint source and channel coding systems.

1 INTRODUCTION

1.1 Motivation

There is a strong similarity between BSS coding relative to Hamming distortion and BSC coding: The same linear block error-correcting code can be used both as a channel code and as a source code, where the channel decoding algorithm is used as a source encoding algorithm, and the channel encoding algorithm is used as a source decoding algorithm. This idea is not new; it appears e.g. in Mc Eliece's book [1, § 11.5].

Such BSS coding techniques—"dual" to BSC coding techniques—have not been widely used, primarily because source encoding requires a *complete* channel decoder. This requires the use of codes with small block lengths so that source encoding can be done using a complete list of syndromes. Thus, practical capabilities of such source coding schemes are limited.

However, our aim is to utilize the similarity between binary source and channel codes to design a *joint* source and channel coding system using *small* linear block codes. Although Shannon's source and channel coding theorem suggests that source and channel coding should be separated, optimum performance would be achieved at the expanse of very large block lengths and very high complexity. In our proposal, complexity is limited by the use of small block lengths yet performance stays relatively close to the optimum.

1.2 Organization of the paper

While the design of error-correcting codes for use on the BSC is well-known, comparatively little is known concerning the use of these codes for BSS data compression with Hamming distortion. For this reason, we derive, in this paper, the optimum performance curve for source coding using block codes for a given block length. For increasing lengths, we can show that these curves converge toward Shannon's limit (known as the rate-distortion function). Exhaustive search for lengths ≤ 11 shows that the best linear codes are always very close to the optimum, for any possible code rate.

Then, we utilize these source and channel codes to design a joint source and channel coding system. We provide simulation results and compare them to Shannon's limit for joint source and channel coding, which is referred to as the optimum performance theoretically attainable (OPTA) in the literature. Although complexity is maintained at a small level, performance is relatively close to the OPTA.

The results given in this paper may easily be extended to q-ary sources, where e.g. q is a prime power integer and codes are defined over a finite field F_q. To simplify the discusssion, we restrict ourselves, in this paper, to the binary case ($q = 2$ and $F_2 = \{0, 1\}$).

2 DUALITY BETWEEN ERROR CORRECTION AND SOURCE CODING

Consider a binary block source coding scheme as illustrated in figure 1. Each m-bit source word $\underline{u}$ is processed to give a k-bit index $\underline{i}$ which represents the source word more compactly (one has $k < m$). The source decoder delivers the m-bit "source codeword" $\underline{v}$ to the user. We

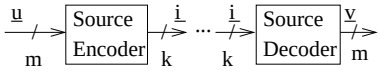

Figure 1: *Block diagram of a binary source coding system*

use the following average distortion per bit as the fidelity criterion.

$$D = \frac{1}{m}\mathbf{E}\left[d_H(\underline{u}, \underline{v})\right] = \frac{1}{m}\sum_{\underline{u}} p(\underline{u}) d_H(\underline{u}, \underline{v}) \qquad (2.1)$$

Here $d_H(\underline{u}, \underline{v})$ is the Hamming distance between the source word and the source codeword (i.e., the number

of places in which they differ), and $p(\underline{u})$ is the probability mass function of the source (for a BSS $p(\underline{u}) = \frac{1}{2^m}$). In other words, D is the average number of bit errors between the system input and output and is, in fact, an error probability citerion.

To each source codeword $\underline{v}$ corresponds a codebook index $\underline{i}$, and, in order to minimize D, $\underline{v}$ should be chosen as the closest codeword to $\underline{u}$ (nearest neighbour condition for the Hamming distance). Therefore, the source encoder should find the codeword that is closest to the source word and then deliver its index $\underline{i}$. The source decoder should perform the inverse 'table lookup' operation to output the corresponding codeword $\underline{v}$.

In this paper, source coding is achieved using a linear (m, k) block code. Such codes were invented to correct errors on noisy channels and were intensively studied (see e.g., [2]). Consider, for example, the BSC coding scheme depicted in figure 2. The channel encoder

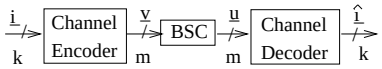

Figure 2: *Block diagram of a binary channel coding system*

adds redundancy to information message $\underline{i}$ and delivers channel codeword $\underline{v}$. The channel decoder uses this redundancy to correct the errors in received word $\underline{u}$; it achieves this by finding the codeword $\underline{v}$ that is closest (in the Hamming sense) to $\underline{u}$, and finally produces an estimate of information message $\underline{i}$.

Therefore, we see that the channel decoding algorithm performs essentially the same task as the source encoding algorithm, which is required for the source coding scheme of figure 1. For this reason, the source encoder will be implemented as a channel decoder. From this follows that we should also use the the channel encoding algorithm as a source decoding algorithm. The code rate $R = \frac{k}{m}$ gives the average number of coded bits in $\underline{i}$ per source bit.

To summarize, binary source and channel coding are essentially dual tasks. There is, however, one important difference concerning the implementation of the source encoder vs. the channel decoder: In many practical channel coding situations, the received word is, with high probability, very close to the emitted channel codeword so that the channel decoder can be *incomplete*. This means that it decodes only those received words having less than a given number of errors. For source coding, however, the input source word can be anywhere. This is especially true for a BSS where source words are equally likely. It is therefore necessary that the source encoder be implemented as a *complete* channel decoder, which produces the closest codeword $\underline{v}$ no matter what the received word $\underline{u}$ is.

Most known efficient channel decoders are incomplete decoders and cannot be used for our purposes. However, complete channel decoding can be performed using the classical syndrome decoding (see e.g. [2]) where finding the closest codeword is done using a complete list of 2^{m-k} syndromes. This, however requires the use of codes with small block lengths in order to maintain complexity at a reasonable level.

3 BEST PERFORMANCE OF BSS CODING

In this section, we derive a closed-form, tight lower bound on the achievable Hamming distortion using source block codes for a given length m and all possible rates $R = \frac{k}{m}$. We then compare this bound to Shannon's rate-distorsion function and show that nearly optimal codes exist for all lengths $m \leq 11$.

3.1 Derivation

Consider an (m, k) binary linear code $\mathcal{C}$ having 2^k codewords $\underline{v}$ of length m. Define the Voronoï cell $V(\underline{v})$ about $\underline{v} \in \mathcal{C}$ as the set of m-bit words that are closest to $\underline{v}$ than to any other codeword. It is easily seen that the Voronoï cells may be chosen so as to be disjoint and to cover the entire space F_2^m.

Now define α_d as the number of words $\underline{u}$ in $V(\underline{v})$ for which $d_H(\underline{u}, \underline{v}) = d$. Since the code is linear, the Voronoï cells are translated copies of each other (and thus enclose exactly 2^{m-k} words). It follows that α_d does not depend on the choice of $V(\underline{v})$. The α_d's are known, in (channel) coding theory [2], as the Hamming weight distribution of the code's coset leaders.

Using these definitions, the Hamming distortion for code $\mathcal{C}$ can easily be rewritten as

$$D = \frac{1}{m2^m} \sum_{\underline{v} \in \mathcal{C}} \sum_{\underline{u} \in V(\underline{v})} d_H(\underline{u}, \underline{v}) = \frac{1}{m2^{m-k}} \sum_{d=0}^{m} d\alpha_d. \quad (3.2)$$

Now pick a $d^* \geq 0$ and split the sum into two parts, one for $d < d^*$ and the other for $d \geq d^*$. In the second sum, use $d \geq d^*$ and the relation $\sum_{d=0}^{m} \alpha_d = |V(\underline{v})| = 2^{m-k}$. We obtain, after re-arranging terms:

$$D \geq \frac{d^*}{m} - \frac{1}{m2^{m-k}} \sum_{d<d^*} (d^* - d)\alpha_d$$

Now use the relation $\alpha_d \leq \binom{m}{d}$, which is easily obtained by counting all words at distance d from a codeword. We end up with the following lower bound on D.

$$D \geq D_{lb}(d^*) = \frac{d^*}{m} - \frac{1}{m2^{m-k}} \sum_{d<d^*} (d^* - d) \binom{m}{d}. \quad (3.3)$$

It can be easily seen that the best lower bound $D_{lb}(d^*)$ is obtained when d^* is the largest integer such that

$$\sum_{d<d^*} \binom{m}{d} \leq 2^{m-k}. \quad (3.4)$$

3.2 Comparison to Shannon's limit

From Shannon's coding theorem, we know that the optimal performance theoretically attainable for BSS coding is given by Shannon's rate-distorsion function

$$R(D) = 1 - H_2(D) \qquad (D \leq \frac{1}{2}) \tag{3.5}$$

where $H_2(x) = -x \log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function. Shannon's initial theorem considers arbitrary (non-linear) codes, but it is known to hold also for the subclass of *linear* codes [3].

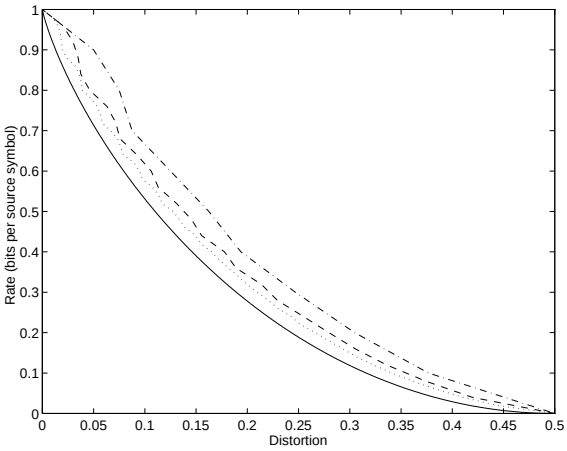

Figure 3: *Lower bound for different block lengths: m=10 (-.), m=25 (–), m=50 (..). Solid: Shannon's rate-distortion function.*

While the rate-distortion function refers to arbitrary large block lengths, our lower bound D_{lb} is established for linear codes of a *fixed* block length m. In figure 3 we have plotted this lower bound as a function of code rate R for different lengths. We observe that the curves $R(D_{lb})$ are always above, and converge toward Shannon's limit $R(D)$ as $m \to \infty$. This can, in fact, be rigorously proved [4].

3.3 In search of good source codes

For short lengths it is possible to carry an exhaustive search of the linear codes achieving the minimum distorsion D for a given rate R. This is done by parameterizing codes with their parity-check matrix in systematic form[1], and computing their standard array [2], which give the coset leader weight distribution $\{\alpha_d\}$, and therefore distortion D by (3.2). The search is limited by the size of the standard array $(2^k \times 2^{m-k})$ and was carried out up to $m = 11$.

Figure 4 gives the performances of the best linear codes for $m = 10$. We have observed that these stand

[1] Without loss of generality, we can restrict the search to systematic codes because two equivalent codes—differing only in the order of bits—clearly achieve the same distortion.

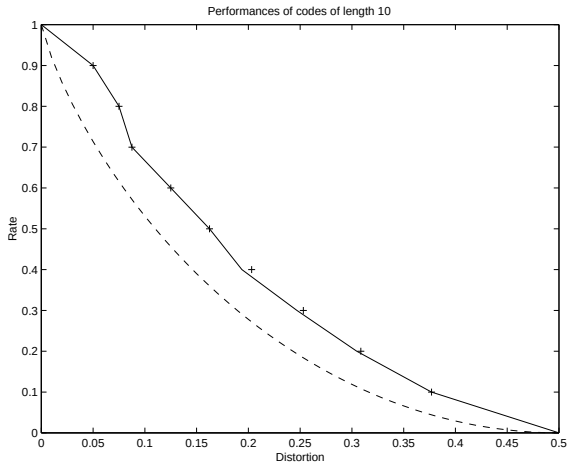

Figure 4: *BSS coding performance for linear codes of length $m = 10$. Solid: $D_{lb}(d)$; dashdotted: Shannon limit $R_s = 1 - H_2(D)$; +: best linear codes found by exhaustive search.*

pretty close to our lower bound D_{lb} for any length $m \leq 11$. In fact, many codes meet the lower bound[2].

4 JOINT SOURCE/CHANNEL CODING

Consider the block diagram of figure 5 for a BSS and a BSC of raw error probability p. Two linear block codes

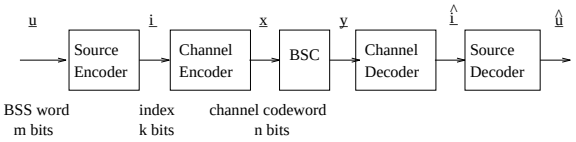

Figure 5: *Block diagram of a source and channel coding system*

are used: One (m, k) code for source coding, and one (n, k) for channel coding. Let $R_s = \frac{k}{m}$ and $R_c = \frac{k}{n}$ be the corresponding source and channel code rates. We define the *global rate* $r = \frac{n}{m} = \frac{R_s}{R_c}$ as the average number of coded bits per source bit. The fidelity criterion is chosen as the global Hamming distortion (2.1), which is, in general, caused not only by the source coder but also by the errors introduced by the channel.

4.1 Optimum performance theoretically attainable (OPTA)

Shannon's theorem on joint source and channel coding [1, chap. 5] states that (1) r and D must satisfy the inequality

$$r \geq \frac{R(D)}{C(p)} = \frac{1 - H_2(D)}{1 - H_2(p)}. \tag{4.6}$$

where $R(D)$ is the BSS rate distorsion function (3.5) and $C(p)$ is the BSC capacity; and (2) that the opti-

[2] We can show [4] that the set of codes meeting the lower bound D_{lb} coincides with the set of perfect and quasi-perfect codes [2].

mum performance $r = \frac{1-H_2(D)}{1-H_2(p)}$ is theoretically achievable. Shannon's proof requires very large block lengths (n and $m \to \infty$) and separate optimization of source and channel coders, where the source encoder and decoder determine the final distortion, and where the channel encoder and decoder achieve a very small error probability which is negligible compared to the distortion introduced by the source encoder. However, by doing the source and channel coding separately, we may end up with a system more complex than necessary.

4.2 Simulation results

We propose a *jointly* optimized coding system in which we restrict ourselves to codes of *small* block lengths—yielding small complexity. For each global rate $r = \frac{n}{m}$, the best combination of source (m, k) and channel (n, k) codes is determined by exhaustiuve search. The results are shown in figure 6. In constrast to Shannon's approach, the final distortion significantly depends on both the source and channel codes, while performances stay relatively close to the OPTA. It is convenient to discuss three cases:

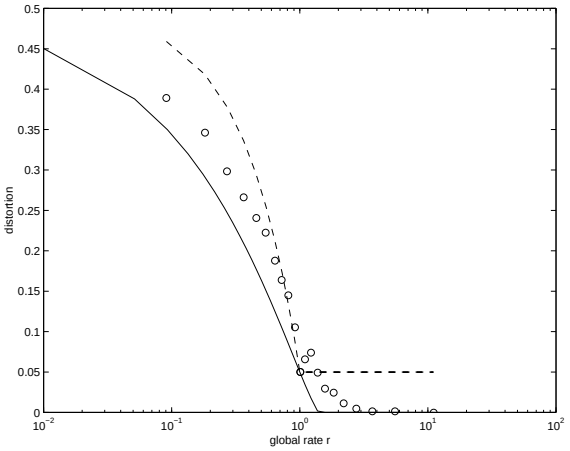

Figure 6: *Distortion D versus global source/channel rate r for a very noisy BSC with raw error probability $p = 0.05$. Solid line: Shannon limit. Dashdotted line: distortion of the 'trivial' systems (see text). Circles: performances of best combination of source (m, k) and channel (n, k) codes for block lengths m and $n \leq 11$.*

Case $r = 1$. From (4.6), the OPTA is $D = p$. It is known (see e.g. [1, Pb. 5.7]), and easily seen, that the OPTA is achieved without any coding at all, by connecting directly the source to the channel[3]. In contrast to Shannon's approach, the resulting system is very simple, and the channel is responsible for the total distortion.

Case $r < 1$. In this case, we observed that the best combination of source and channel coders occurs when the channel code is a universal code—that is, $k = n$. Thus, in this case, the best way to do joint coding is to do no channel coding at all. We also have an empirical argument [4], similar to the one derived for $r = 1$, which suggests that this should be so. Figure 6 shows the performances of the best source codes as derived in 3.3 for $m \leq 11$. It also makes the comparison to 'trivial' codes for which only the fraction r of the source bits are transmitted and the source decoder guesses the rest by flipping an unbiased coin, yielding distortion $D = pr + \frac{1}{2}(1-r)$. Observe that the trivial codes perform better than the 'best' source codes when r is close to 1.

Case $r > 1$. In this case, we observed that the best combination of source and channel coders occurs when the source code is a universal code—that is, $m = k$. Thus, in this case, the best way to to do joint source/channel coding is to do no source coding at all. Figure 6 shows the performances of the best channel codes for $n \leq 11$, which were derived in a manner similar to what was done in 3.3. It also makes the comparison to 'trivial' codes for which the m-bit source word $\underline{u}$ is padded with $n - m$ zeros to give channel codeword $\underline{x}$, yielding distortion $D = p$. Observe that the trivial codes perform better than the 'best' source codes when r is close to 1.

[3] Moreover, we can show by calculation [4] that this trivial system is the only combination of source and channel coders for which OPTA is attained.

5 Conclusion

In this paper, we have first discussed the strong similarity between binary symmetric source coding and binary symmetric channel coding: They are essentially dual tasks.

We have then utilized this duality for designing data compression systems with error probability criterion, and have derived the optimum performance curve using block codes of a given length. For increasing lengths, these curves converge toward Shannon's limit. Exhaustive search for lengths ≤ 11 shows that the best linear codes are always very close to the optimum.

Finally, these source codes were used to construct simple joint source and channel coding systems. In our proposal, complexity is limited by the use of small block lengths yet performance stays relatively close to the optimum performance theoretically attainable.

References

[1] McEliece R.J., *The theory of information and coding*, Addison Wesley, Reading, MA, 1977.

[2] F.J. MacWilliams, N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, 1977.

[3] Goblick T.J., *Coding for a discrete information source with a distortion measure*, Ph.D., MIT, Cambridge, MA, 1962.

[4] Bergot F.X. and Rioul O., in preparation.