

Superposition Coding for Source Coding with Side Information at the Decoder

Lorenzo Cappellari and Giancarlo Calvagno

Dept. of Information Engineering, University of Padova, Italy

Email: {lorenzo.cappellari, giancarlo.calvagno}@dei.unipd.it

Abstract—Problems of noisy channel coding with side information at the encoder and of lossy source coding with side information at the decoder have many practical applications. Unfortunately, the theoretical performance bounds are obtained by randomly partitioning a randomly generated source or channel code into an ensemble of channel or source codes, which is highly non-practical. A scheme has recently appeared for the problem of channel coding with side information where two random but independently generated codes lead to achievement of the cost-capacity function. Moreover, in practice, in this scheme the theoretical bounds are also achieved by using practical, more structured codes. In this paper, we investigate a similar approach for the problem of source coding with side information. We show that when the virtual correlation between the side information and the source is additive, a source encoder based on a codebook obtained as the superposition of two independent random codes achieves the rate-distortion function with side information. Finally, we discuss the challenges related with the practical application of this scheme.

I. INTRODUCTION

The problem of source coding in presence of some correlated side information at the decoder which is not accessible by the encoder has many applications. In particular, it arises in all sensor networks, where the power for communication is severely constrained [1], and in sequential source coding of correlated data when for some reasons (i.e. encoding complexity or robustness against transmission errors) data must be encoded separately, as in video coding [2].

In the lossless case, the minimum achievable transmission rate was found by Slepian and Wolf, who showed that there is no rate penalty with respect to the case in which the side information is also available at the encoder [3]. In the lossy case, where the goal is to minimize the transmission rate under a given distortion constraint on the reconstructed source, the *rate-distortion function* was found by Wyner and Ziv [4]. In the latter case, in general, a rate penalty may occur in having the side information only at the decoder. In both cases, the achievability of these rate bounds was shown using randomly generated codes whose codewords are randomly grouped into bins.

In practice, codes having some structure must be used. In the Slepian-Wolf problem (lossless/near-lossless case) all “good” practical codes currently employed in channel coding (such as turbo [5] and LDPC [6] codes) induce a partition of the set of all source outcomes that represents in fact a very good binning. Hence, the utilization of these codes leads to practical schemes

with performances close to the theoretical bounds [7]. Similar results are also obtained in lossless *multiterminal* network coding [8]. In the Wyner-Ziv problem (lossy coding) practical code design is more difficult. Zamir et al. showed that *nested linear codes* (in the binary case) or *nested lattice codes* (in the continuous case) achieve asymptotically the rate-distortion bound [9]. Similarly, the same structures are shown to achieve asymptotically the *capacity-cost function* in the channel coding problem where side information regarding the channel state is non-causally available at the encoder but not at the decoder [10]. But, in practice, *finite-dimensional* linear/lattice codes must be used as proposed in [11] or [12] that limit the achievable performances. For this reason, in [12] a scheme is proposed where finite-dimensional *nested quantization* on nested lattices is followed by a second Slepian-Wolf “binning” stage in order to limit the performance loss. At high rates, it is shown that the scheme performs as conventional entropy-coded lattice quantization with side information known as well at the encoder. Similar solutions can be employed also in general *multiterminal* source code design [13].

The aim of this paper is to propose and investigate a simpler scheme for Wyner-Ziv coding, in which the theoretical performance could be achieved using two *independent* codes, “good” for source and channel coding purposes respectively. Practical codes for the source and the channel coding problems are in fact nowadays well-known, while nested structures have been essentially studied only from a theoretical perspective.

For the case of channel coding with additive noise and an additive *interference* signal non-causally known at the encoder but not at the decoder (i.e. the so-called *dirty paper* problem [14]), a practical scheme has been recently proposed [15] which achieves the capacity-cost function using two *independent* rather than nested codes. In particular, the original problem is transformed into an equivalent *multiple-access* problem in which the *superposition* of the two codes is decoded using successive interference cancellation. Very good performances are shown for both the binary [16] and the Gaussian [17] setting. Since the codes can be independently designed, each one of them can be specifically tailored for the purpose it serves. In fact, one of them mainly serves as source code (on which we must be able to perform *quantization*), while the other serves as channel code (on which we must be able to apply some channel-decoding operations with performance close to the one of *joint-typicality* used in random analysis).

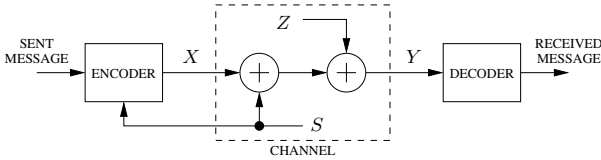


Fig. 1. Channel coding with side information.

Since the problem of lossy source coding with side information is “dual” to the problem of noisy channel coding with side information, it may seem straightforward that a similar superposition approach with independent codes may be effectively used in the Wyner-Ziv setting too. However, the encoder/decoder of one problem *functionally* corresponds to the decoder/encoder of the other one only under certain hypothesis [18]. For example, in the binary setting, the two problems are not exactly duals in the sense discussed in [18]. Moreover, functional duality assumes that both encoder and decoder operate an exact joint-typicality operation, which in practice is not the case. For example, belief-propagation in traditional channel decoding of turbo/LDPC codes roughly corresponds to joint-typicality only if the input to the decoder is “close” to an actual codeword, but not in general. Hence, in this paper, we aim to analyze the superposition approach for the Wyner-Ziv problem without relying on duality.

The rest of this paper is organized as follows. In Section II we review the superposition coding approach for the dirty paper problem. In Section III we analyze the performance of superposition coding for source coding with side information from a random coding perspective. The result of this analysis is then particularized for the binary and the Gaussian case. The issues involved in a practical implementation of this method are discussed in Section IV. Section V summarizes our conclusions.

II. SUPERPOSITION CODING FOR WRITING ON DIRTY PAPER

Consider the additive memoryless channel

$$Y = X + S + Z,$$

where X is the input to the channel, S is an interference signal known (non-causally) to the encoder and independent from X , Z is an unknown channel noise independent from X and S , and Y is the channel output. Assume that the channel is used n times, without feedback; X^n , S^n , Z^n , and Y^n denote the involved random vectors. The channel is graphically shown in Fig. 1, and the goal is to achieve its capacity, i.e. to transfer the maximum amount of information per channel use to the receiver, given a certain cost constraint on X .

The general expression of the capacity-cost function for channel coding with side information (for a known distribution $p(s)$ and channel transfer function $p(y|x,s)$) was found in [10], and for a cost function $w(\cdot)$ and a cost constraint W it equals

$$C(W) = \max_{p(u|s), p(x|u,s)} I(U; Y) - I(U; S), \quad (1)$$

where $I(\cdot; \cdot)$ denotes mutual information, U is an auxiliary random variable such that $Y \rightarrow (X, S) \rightarrow U$ and $Y \rightarrow (U, S) \rightarrow X$ are Markov chains [19], and $E[w(X)] = \sum_{s,u,x} p(s)p(u|s)p(x|u,s)w(x) \leq W$.

In the binary case the alphabet over which the random variables take values is $\mathcal{A} = GF(2)$, “+” indicates the sum over the field $GF(2)$, $S \sim \mathcal{B}(1/2)$ (i.e. is distributed as Bernoulli-1/2), $Z \sim \mathcal{B}(p)$, and X is subject to the cost constraint $E[d_H(X^n, 0^n)/n] \leq W$, where $d_H(\cdot, \cdot)$ is the Hamming distance. Under these conditions, (1) can be exactly evaluated and becomes

$$C(W) = \begin{cases} W \log_2 \frac{1-W_c}{W_c} & \text{for } 0 \leq W < W_c \\ H(W) - H(p) & \text{for } W_c \leq W \leq 1/2 \end{cases}, \quad (2)$$

where $H(q)$ denotes the entropy of a random variable $\sim \mathcal{B}(q)$, and $W_c = 1 - 2^{-H(p)}$.

Two codes $\mathcal{C}_0 \subset \mathcal{A}^n$ and $\mathcal{C}_1 \subset \mathcal{A}^n$ are constructed at rates R_0 and R_1 (bit/symbol) by random i.i.d. selection according to distributions $\mathcal{B}(1/2)$ and $\mathcal{B}(q)$, respectively.

Given x^n and a code \mathcal{C} , let $T_{\mathcal{C}}(x^n)$ denote a codeword of \mathcal{C} which is (strongly) jointly typical with x^n , if there exist at least one; otherwise, let it denote a random codeword of \mathcal{C} . Define $[x^n]_{\mathcal{C}} \triangleq T_{\mathcal{C}}(x^n) - x^n$. The encoder selects a codeword $c_1^n \in \mathcal{C}_1$ and sends the sequence

$$x^n = [s^n - c_1^n]_{\mathcal{C}_0}.$$

If $T_{\mathcal{C}_0}(s^n - c_1^n) \triangleq c_0^n \in \mathcal{C}_0$, the channel output is

$$y^n = c_0^n - (s^n - c_1^n) + s^n + z^n = c_0^n + c_1^n + z^n.$$

The decoder computes the pair $(\hat{c}_0^n, \hat{c}_1^n) \in \mathcal{C}_0 \times \mathcal{C}_1$ such that it is jointly typical with y^n , announcing \hat{c}_1^n as the decoded message.

By random analysis over all possible tuples of codes (see [15]), if $R_0 > R_{1/2}(W)$, where $R_{1/2}(W) = 1 - H(W)$ is the Hamming-distortion rate-distortion function [19] of a binary symmetric source, the average probability of having X^n violate the cost constraint W (encoder error) approaches zero with $n \rightarrow \infty$.¹ In addition, if $R_1 \leq C_p(q)$ and $R_0 + R_1 \leq C_p(1/2)$, where $C_p(q) = H(p*q) - H(p)$ ($p*q \triangleq p(1-q) + q(1-p)$) is the Hamming-weight cost-capacity function of a binary symmetric channel with error probability p , the probability of having $(\hat{c}_0^n, \hat{c}_1^n) \neq (c_0^n, c_1^n)$ (decoder error) vanishes with $n \rightarrow \infty$.² If a q^* is chosen such that $p*q^* = W$, then reliable transmission (i.e. without encoder nor decoder error) is possible at rates arbitrarily close to $R_1^* = H(W) - H(p)$, which equals the cost-capacity function with side information at the encoder of the considered channel (2), at least for all constraints $W_c \leq W \leq 1/2$.

In the continuous case the alphabet over which the random variables take values is $\mathcal{A} = [-A/2, A/2] \sim \mathbb{R}/AZ$ (for some $A > 0$), “+” indicates the modulo- A sum, $Z \sim \mathcal{N}_A(0, P_Z)$ (i.e. is distributed as A -aliased Gaussian variable with zero

¹Note that in fact $S^n - C_1^n$ is a binary symmetric source.

²Note that in fact $C_0^n + C_1^n + Z^n$ defines a binary symmetric channel in both the case with the codewords of \mathcal{C}_1 as input (and known C_0^n) or the case with codewords of $\mathcal{C}_0 + \mathcal{C}_1$ as input.

mean and variance P_Z), and X is subject to the cost constraint $E[\|X^n\|^2/n] \leq P_X$, where $\|\cdot\|^2$ is the square-distance norm. If $A \rightarrow \infty$, (1) can be exactly evaluated and becomes

$$C(P_X) = \frac{1}{2} \log_2 \left(1 + \frac{P_X}{P_Z} \right), \quad (3)$$

i.e. equals the cost-capacity function of the case without side information.

The two codes $\mathcal{C}_0 \subset \mathcal{A}^n$ and $\mathcal{C}_1 \subset \mathcal{A}^n$ are constructed at rates R_0 and R_1 (bit/symbol) by random i.i.d. selection according to distributions $\mathcal{U}[-A/2, A/2)$ (i.e. uniform) and $\mathcal{N}_A(0, Q)$, respectively.

Assume that D is a *dither* signal known to both encoder and decoder, drawn accordingly to $\mathcal{U}[-A/2, A/2)$, and independent from all other variables. The encoder selects a codeword $c_1^n \in \mathcal{C}_1$ and sends the sequence

$$x^n = [\alpha s^n - c_1^n + d^n]_{\mathcal{C}_0}.$$

If $T_{\mathcal{C}_0}(\alpha s^n - c_1^n + d^n) \triangleq c_0^n \in \mathcal{C}_0$, the channel output is

$$y^n = c_0^n - (\alpha s^n - c_1^n + d^n) + s^n + z^n.$$

The decoder first computes

$$\hat{y}^n = \alpha y^n + d^n = c_0^n + c_1^n + \underbrace{[\alpha z^n - (1 - \alpha)x^n]}_{\hat{z}^n},$$

and then finds the pair $(\hat{c}_0^n, \hat{c}_1^n) \in \mathcal{C}_0 \times \mathcal{C}_1$ such that it is jointly typical with \hat{y}^n , announcing \hat{c}_1^n as the decoded message. $\alpha = P_X/(P_X + P_Z)$ minimizes the power of the equivalent noise in this virtual multiple-access channel (MAC), producing $P_{\hat{Z}} = \alpha P_Z$.

Again, if in \mathcal{C}_0 there are enough codewords, it is possible to have a vanishing probability of encoding errors; in addition, if in \mathcal{C}_1 and in $\mathcal{C}_0 + \mathcal{C}_1$ there are not too many codewords, it is possible to have a vanishing probability of decoding errors. In particular when $A \rightarrow \infty$ (i.e. when the channel is AWGN), if a Q^* is chosen such that $P_{\hat{Z}} + Q^* = P_X$, then reliable transmission (i.e. without encoder nor decoder error) is possible at rates arbitrarily close to $R_1^* = (1/2) \log_2(1 + P_X/P_Z)$, which equals the cost-capacity function with side information at the encoder (3) of the considered channel [14]. Details are given in [15].

Note that in both cases the two codes take a very different role. The code \mathcal{C}_1 must be designed in order to be a good channel code with respect to the virtual MAC arising in the decoding operation. The code \mathcal{C}_0 has instead dual requirements to be both a good source code, in order to avoid the encoder error, and a good channel code, for effective MAC decoding. However, since we are not interested in exactly decoding the correct c_0^n , one should mainly be concerned in choosing a code \mathcal{C}_0 which is good for source coding purposes. The same exact considerations arise when analyzing this superposition scheme from a practical perspective. In practice, in source encoding the joint-typicality operation is replaced by maximum likelihood decoding, and structured codes which allow for exhaustive search over all codewords (e.g. trellis codes) are employed;

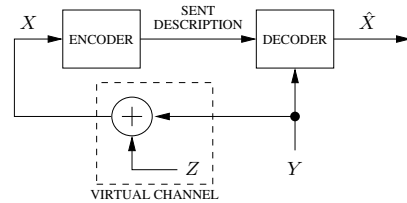


Fig. 2. Source coding with side information.

in channel decoding the joint-typicality operation is approximated using some belief propagation algorithm conducted on codes with a higher degree of randomness (e.g. turbo or LDPC codes). For these reasons, in practical schemes where \mathcal{C}_0 is a trellis code, \mathcal{C}_1 is a turbo code, and iterative algorithms are employed during decoding, the performances are very close to the theoretical bounds, even if \mathcal{C}_0 is not good from a channel coding perspective.

III. SUPERPOSITION CODING FOR WYNER-ZIV CODING

The superposition approach for writing on dirty paper is feasible because of the additive effect of the interference signal (known at the encoder). Let us consider a tuple of correlated random variables (X, Y) such that

$$X = Y + Z,$$

for some Z independent from Y ; X is the input to the encoder and Y is the side information (known at the decoder); \hat{X} denotes the reconstruction of X at the decoder. The alphabet over which these random variables take values is a *finite group* \mathcal{G} with 2^l elements ($l \geq 1$), “+” indicates the sum over this group, $Y \sim \mathcal{U}(\mathcal{G})$ (i.e. is uniformly distributed over \mathcal{G}), $Z \sim p(z)$, and \hat{X} is subject to the distortion constraint $E[d_H(X^n, \hat{X}^n)/n] \leq D$. The coding scheme is graphically shown in Fig. 2, and the goal is to achieve the rate-distortion function, i.e. to transfer the minimum amount of information from the encoder to the decoder, given the distortion constraint between \hat{X} and X .

The general expression of the rate-distortion function for source coding with side information (for a known distribution $p(x, y)$) was found in [4], and for a distortion function $d(\cdot, \cdot)$ and a distortion constraint D it equals

$$R(D) = \min_{p(u|x), p(\hat{x}|u, y)} I(U; X) - I(U; Y), \quad (4)$$

where U is an auxiliary random variable such that $Y \rightarrow X \rightarrow U$ and $X \rightarrow (U, Y) \rightarrow \hat{X}$ are Markov chains [19], and $E[d(\hat{X}, X)] = \sum_{x, y, u, \hat{x}} p(x, y) p(u|x) p(\hat{x}|u, y) d(\hat{x}, x) \leq D$.

The two codes $\mathcal{C}_0 \subset \mathcal{G}^n$ and $\mathcal{C}_1 \subset \mathcal{G}^n$ are constructed at rates R_0 and R_1 (bit/symbol) by random i.i.d. selection according to distributions $\mathcal{U}(\mathcal{G})$ and $q(\cdot)$, respectively.

The encoder looks for a codeword $(c_0^n + c_1^n)$, with $c_i^n \in \mathcal{C}_i$, such that it is (strongly) jointly typical with x^n , and sends c_1^n to the decoder (if there are no jointly typical codewords, a random codeword is sent). There is an encoder error if $d_H(x^n, c_0^n + c_1^n) > D$.

Theorem 1: The probability of having an encoder error vanishes with $n \rightarrow \infty$ if

$$\begin{aligned} R_0 &> l - H(q * d) \\ R_0 + R_1 &> R_{\mathcal{U}}(D), \end{aligned}$$

where $*$ denotes probability mass convolution, $H(q)$ denotes the entropy of a random variable $\sim q(\cdot)$, $d(\cdot)$ is the distribution which takes zero with probability $1 - D$ and all other symbols with probability $\frac{D}{2^i - 1}$, and $R_{\mathcal{U}}(D) = l - H(d)$ is the Hamming-distortion rate-distortion function of the uniform random variable³.

Proof: Asymptotically, x^n takes values on a set of $2^{nH(X)}$ elements, and for a fixed c_0^n , the code \mathcal{C}_1 covers at most $2^{nH(q*d)}$ of them within distortion D . Hence, there must be at least $2^{nH(X) - nH(q*d)}$ codewords in code \mathcal{C}_0 . Then, since there are at most $2^{nH(d)}$ covered elements within *balls* of distortion D , the two codes must provide at least $2^{nH(X) - nH(d)} = 2^{nR_{\mathcal{U}}(D)}$ codewords. Once we have these two conditions the probability of finding a typical $(c_0^n + c_1^n)$ approaches one and the distortion constraint is not violated. This may be proved with an argument similar to the one used in the standard proof of the achievability of the rate-distortion function [19].

The decoder receives the codeword $c_1^n = x^n - c_0^n + \hat{z}^n$, where \hat{Z} is independent from X and distributed as $d(\hat{z})$ (it corresponds to the *subtractive* noise in the equivalent symmetric *test channel* between \hat{X} and X). Then it computes

$$[y^n - c_1^n]_{\mathcal{C}_0} = [c_0^n - (z^n + \hat{z}^n)]_{\mathcal{C}_0} \stackrel{c.d.}{=} z^n + \hat{z}^n \quad (5)$$

and finally reconstructs

$$\hat{x}^n = y^n + [y^n - c_1^n]_{\mathcal{C}_0} \stackrel{c.d.}{=} x^n + \hat{z}^n.$$

Equality in (5) is conditional on correct decoding (no decoding errors), i.e. it holds only if there is only one codeword in \mathcal{C}_0 jointly typical with $(y^n - c_1^n)$.

Theorem 2: The probability of having a decoder error vanishes with $n \rightarrow \infty$ if

$$R_0 \leq C_{p*d},$$

where $C_{p*d} = l - H(p*d)$ is the (unconstrained) capacity of an additive channel on \mathcal{G} with noise distributed as $p*d(\cdot)$.

Proof: Asymptotically, the equivalent noise $(z^n + \hat{z}^n)$ is distributed according to $p*d(\cdot)$ (note that \hat{Z} is independent from Z), and hence takes values on a set having $2^{nH(p*d)}$ elements. Hence, there can be at most $2^{nH(X) - nH(p*d)} = 2^{nC_{p*d}}$ non overlapping codewords in \mathcal{C}_0 . Again, once this condition is met the achievability of no decoding errors may be proved with an argument similar to the one used in the standard proof of the achievability of channel capacity [19].

The rate region where there are no errors is not empty for any $q(\cdot)$ such that $H(q*d) > H(p*d)$ and is shown in Fig. 3. In the following, we will particularize this result for the doubly symmetric binary and for the Gaussian case, and show that the Wyner-Ziv bound can be achieved in both cases.

³ $d(\cdot)$ maximizes $H(d)$ and $H(q*d)$ over all distributions with $d(0) = 1 - D$.

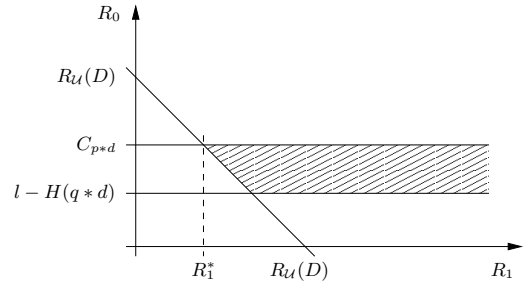


Fig. 3. Region in which there are no encoder nor decoder errors.

A. Binary Sources

In case of $\mathcal{G} = GF(2)$, where $Z \sim \mathcal{B}(p)$ and $C_1 \sim \mathcal{B}(q)$, the rate-distortion function and the channel capacity involved in the calculation of the rate region equal

$$\begin{aligned} R_{\mathcal{U}}(D) &= 1 - H(D) \\ C_{p*d} &= 1 - H(p*D), \end{aligned}$$

respectively. Hence, the lowest achievable value for R_1 is

$$R_1^* = R_{\mathcal{U}}(D) - C_{p*d} = H(p*D) - H(D),$$

that is the rate-distortion function with side information for all distortions $0 \leq D \leq D' < p$ [4]. The rate-distortion function for $D' < D \leq p$ is achieved by time-sharing of the two working points $(H(p*D') - H(D'), D')$ and $(0, p)$.⁴

Under these conditions, in fact, (4) can be exactly evaluated and becomes

$$R(D) = \begin{cases} H(p*D) - H(D) & \text{for } 0 \leq D \leq D' \\ R(D') \frac{p-D}{p-D'} & \text{for } D' < D \leq p \end{cases} \quad (6)$$

In particular, for $0 \leq D \leq D'$ the minimum in (4) is achieved by $U = X + Q$ and $\hat{X} = U$, where $Q \sim \mathcal{B}(D)$ is independent from X .

B. Gaussian Sources

Assume now that $\mathcal{G} = \mathbb{R}$, $Y \sim \mathcal{N}(0, P_Y)$, $Z \sim \mathcal{N}(0, P_Z)$, $C_0 \sim \mathcal{N}(0, P_0)$, and $C_1 \sim \mathcal{N}(0, Q)$. The constraint is given in terms of the squared Euclidean distance. In order to analyze the Gaussian case, which is not discrete, some care should be taken because of the fact that while the *test channel* between \hat{X} and X is still additive, the equivalent channel between X and \hat{X} is additive only if a suitable scaling factor is introduced [19].

In this case, assuming that $U \sim \mathcal{N}(0, D)$ is a *dither* signal known to both encoder and decoder, and independent from all other variables, the encoder sends the c_1^n corresponding to $(\beta x^n + u^n)$, i.e. the decoder receives

$$c_1^n = \beta x^n + u^n - c_0^n + \hat{z}^n,$$

where \hat{Z} is (a scaled version of) the additive noise in the channel from X to \hat{X} , and is independent from X (and from Z). If the rate R_0 and the sum $R_0 + R_1$ are high enough, there

⁴ D' is such that $\left. \frac{d[H(p*D) - H(D)]}{dD} \right|_{D'} = \frac{H(p*D') - H(D')}{D' - p}$.

exist codes such that \hat{Z} has maximum power D (no encoder error).

The decoder evaluates

$$[\beta y^n + u^n - c_1^n]_{\mathcal{C}_0} = [c_0^n - (\beta z^n + \hat{z}^n)]_{\mathcal{C}_0} \stackrel{c.d.}{=} \beta z^n + \hat{z}^n \quad (7)$$

and finally reconstructs

$$\hat{x}^n = y^n + \beta[\beta y^n + u^n - c_1^n]_{\mathcal{C}_0} \stackrel{c.d.}{=} x^n + (\beta \hat{z}^n - (1 - \beta^2)z^n).$$

If $\beta = \sqrt{1 - D/P_Z}$, the power of $(\beta \hat{z}^n - (1 - \beta^2)z^n)$ is minimized and equals exactly D ; the power of $(\beta z^n + \hat{z}^n)$ equals P_Z . Hence, if the rate R_0 is less than a certain threshold, we can have correct decoding in (7), i.e. no decoding error. The minimum achievable rate R_1^* in this case can be asymptotically computed with a geometric argument: the final goal is to cover each ball related to a codeword of \mathcal{C}_0 (which has at least power P_Z) with as least as possible balls of power D , each one of them related to one codeword of \mathcal{C}_1 (for which $Q + D > P_Z$). Finally we obtain

$$R_1^* = \frac{1}{2} \log_2 \left(\frac{P_Z}{D} \right),$$

which equals the rate-distortion function with side information in the Gaussian case.

Under this condition, in fact, (4) can be evaluated and leads exactly to this result. In particular, the minimum in (4) is achieved by $U = \beta^2 X + Q$ and $\hat{X} = (1 - \beta^2)Y + U$, where $Q \sim \mathcal{N}(0, \beta^2 D)$ is independent from X .

The power P_0 must be enough in order to have the codewords of $\mathcal{C}_0 + \mathcal{C}_1$ cover all the space in which asymptotically $\beta x^n + u^n$ lies.

IV. IMPLEMENTATION ISSUES

As clear from the previous section, with superposition coding it is possible to find two independent codes that guarantee the achievability of the Wyner-Ziv bound in the two examined cases. In particular, the code \mathcal{C}_1 must offer a good *covering* of the space, i.e. it must be good for source coding purposes. On the other side, \mathcal{C}_0 must take the role of both a good code for source coding and a good code for channel coding. However, even if \mathcal{C}_0 was not very good from a source coding perspective, it is reasonable that increasing the *power* (q or Q) of \mathcal{C}_1 the superposed code is still good for source coding. It is instead crucial that \mathcal{C}_0 is a good channel code in order to avoid the decoding errors.

The best source codes available today are represented by the trellis codes [20] which offer performance very close to the rate-distortion function. For this codes, given a random realization of the variable to be quantized, it seems crucial that there is the possibility to search for the closest (i.e. the most likely) codeword by examining all codewords. Hence, \mathcal{C}_1 should in general be a good trellis code.

The best channel codes are instead represented by turbo and low-density parity-check codes (LDPC). Those codes have a higher degree of randomness, which prevents (from a computational complexity point of view) conducting the search over all codewords for applying exactly a maximum likelihood

approach. However, there exist very good *message passing algorithms* over their *factor-graphs* [21], which almost always converge to the most likely output. Then, \mathcal{C}_0 should be one of these codes.

Unfortunately, the message-passing algorithms fail to converge when the input distribution is not unimodal and centered over an actual codeword, as it happens in quantization. In the superposition coding approach, in principle, we should be able to quantize the source outcome over \mathcal{C}_0 and successively quantize a *residual* over \mathcal{C}_1 (this would be similar to successive interference cancellation used in the MAC case). For the reason mentioned above, the currently available algorithms do not allow to perform the first quantization.

An approach in which \mathcal{C}_0 is a convolutional code was presented in [22] and achieved a 3 ÷ 4 dB gap with respect to the Wyner-Ziv bound. That gap is due to the fact that the performance of convolutional codes is somewhat far from the channel capacity. Nevertheless, the scheme turned out to have a very simple implementation, and was shown to have a great degree of adaptation both to the source-side correlation and to the desired transmission rate. For these reasons, it was also tested in the context of video coding [23], [24]. In addition, if \mathcal{C}_1 is an embedded source code such that it provides a means of scalable quantization, it can be shown that also the resulting Wyner-Ziv coding system inherits the scalability property with an acceptable performance. Nowadays, scalable source coding has been deeply investigated, so that practical schemes can be easily included in the superposition framework outlined in this paper.

For practical implementation of the superposition approach, which still remains interesting for the possibility to use *independent* codes, it will be necessary to find good sparse codes and develop good algorithms for lossy quantization over them. One of the first works showing that iterative algorithms may work for quantization too appeared in [25]; that result was obtained by duality with good codes for binary erasure channels. More recently, schemes based on *survey propagation* were adapted from the field of statistical physics in order to do data compression as well [26]. Among the sparse codes that are currently under investigation for source coding, there is an high interest in low-density generator matrix codes (LDGM) which are shown to achieve the rate-distortion bound [27], [28]. Nevertheless, general algorithms which allow for practical utilization of these codes have not appeared yet. Other somewhat more practical approaches have appeared in [29], [30], [31].

V. CONCLUSION

In this paper we discussed the superposition coding approach for the problem of source coding with side information at the decoder. For the case of a general additive-symmetric discrete *correlation* channel between the (uniform) side information and the source, we derived a rate region for the two independent superposed codes in which the desired distortion bound can be achieved. We showed that in the binary case the Wyner-Ziv bound is achievable, and extended the same

result to the Gaussian case with Gaussian side information. Finally, we discussed the implementation issues involved in this scheme, which requires quantization over a code which must be a “good” code from a channel coding perspective.

REFERENCES

- [1] Z. Xiong, A. D. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 80–94, Sep. 2004.
- [2] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed video coding,” *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.
- [3] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [4] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [5] C. Berrou and A. Glavieux, “Near optimum error correcting coding and decoding: turbo-codes,” *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.
- [6] D. MacKay, “Good error-correcting codes based on very sparse matrices,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
- [7] J. Garcia-Frias, Y. Zhao, and W. Zhong, “Turbo-like codes for transmission of correlated sources over noisy channels,” *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 58–66, Sep. 2007.
- [8] V. Stanković, A. D. Liveris, Z. Xiong, and C. N. Georghiades, “On code design for the Slepian-Wolf problem and lossless multiterminal networks,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1495–1507, Apr. 2006.
- [9] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.
- [10] S. Gel'fand and M. Pinsker, “Coding for channel with random parameters,” *Probl. Contr. Inf. Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.
- [11] S. D. Servetto, “Lattice quantization with side information: Codes, asymptotics, and applications in sensor networks,” *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 714–731, Feb. 2007.
- [12] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, “Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design,” *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4358–4379, Oct. 2006.
- [13] Y. Yang, V. Stanković, Z. Xiong, and W. Zhao, “On multiterminal source code design,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2278–2302, May 2008.
- [14] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [15] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, “Superposition coding for side-information channels,” *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1872–1889, May 2006.
- [16] G. Caire, A. Bennatan, D. Burshtein, and S. Shamai, “Coding schemes for the binary symmetric channel with known interference,” in *Proc. of the 41st annual Allerton Conf. on Commun. Control and Computing*, vol. 41, Oct. 2003, pp. 809–818.
- [17] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, “Superposition coding for Costa channels,” in *Proc. of IEEE Intl. Symp. on Inf. Theory and its Applications*, Oct. 2004.
- [18] S. S. Pradhan, J. Chou, and K. Ramchandran, “Duality between source coding and channel coding and its extension to the side information case,” *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.
- [20] M. W. Marcellin and T. R. Fisher, “Trellis coded quantization of memoryless and Gauss-Markov sources,” *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, Jan. 1990.
- [21] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [22] L. Cappellari and G. A. Mian, “A practical algorithm for distributed source coding based on continuous-valued syndromes,” in *Proc. of European Signal Process. Conf. (EUSIPCO)*, Florence, Italy, 4–8 Sep. 2006.
- [23] —, “An algorithm for intra-frame video coding based on continuous-valued syndromes,” in *Conf. Rec. of 40th IEEE Asilomar Conf. on Signals, Syst. and Comput.*, Pacific Grove, CA, U.S.A., 29 Oct.–1 Nov. 2006, pp. 1090–1094.
- [24] L. Cappellari, “Wavelet-domain distributed video coding based on continuous-valued syndromes,” in *Proc. of European Signal Process. Conf. (EUSIPCO)*, Poznań, Poland, 3–7 Sep. 2007, pp. 1422–1426.
- [25] E. Martinian and J. S. Yedidia, “Iterative quantization using codes on graphs,” in *Proc. of 41st Annual Allerton Conf. on Commun., Control and Comput.*, Oct. 2003, pp. 1317–1326.
- [26] S. Ciliberti, M. Mezard, and R. Zecchina, “Message passing algorithms for non-linear nodes and data compression,” Aug. 2005. [Online]. Available: <http://arxiv.org/pdf/cond-mat/0508723>
- [27] E. Martinian and M. Wainwright, “Low density codes achieve the rate-distortion bound,” in *Proc. of IEEE Data Compression Conf.*, Mar. 2006, pp. 153–162.
- [28] M. J. Wainwright, “Sparse graph codes for side information and binning,” *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 47–57, Sep. 2007.
- [29] M. J. Wainwright and E. Maneva, “Lossy source encoding via message-passing and decimation over generalized codewords of LDGM codes,” in *Proc. of IEEE Intl. Symp. on Inf. Theory*, Sep. 2005, pp. 1493–1497.
- [30] A. Gupta and S. Verdú, “Nonlinear sparse-graph codes for lossy compression,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 1961–1975, May 2009.
- [31] T. Filler and J. Fridrich, “Binary quantization using belief propagation with decimation over factor graphs of LDGM codes,” in *Proc. of 45th Annual Allerton Conf. on Commun., Control and Comput.*, Sep. 2007, pp. 495–501.