Practical distributed source coding with impulse-noise degraded side information at the decoder

Claudio Weidmann, Francesca Bassi, Michel Kieffer

PRACTICAL DISTRIBUTED SOURCE CODING WITH IMPULSE-NOISE DEGRADED SIDE INFORMATION AT THE DECODER

Claudio Weidmann,[1] *Francesca Bassi,*[2] *and Michel Kieffer*[2]

[1]ftw. Telecommunications Research Center, Donau-City-Strasse 1, A-1220 Vienna, Austria
[2]LSS – CNRS – SUPELEC – Univ Paris-Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France

ABSTRACT

This paper introduces a practical method for distributed lossy compression (Wyner-Ziv quantization) with side information available only at the decoder, where the side information is equal to the signal affected by background noise and additional impulse noise. At the core of the method is an LDPC-based lossless distributed (Slepian-Wolf) source code for q-ary alphabets, which is matched to the impulse probability and allows to remove the scalar-quantized impulse noise. Applications of this method to distributed compressed sensing of signals that differ in a sparse set of locations is also discussed, as well as some differences and similarities of variable- and fixed-length coding of sparse signals.

1. INTRODUCTION

Distributed source coding is receiving a lot of attention for a variety of applications, such as distributed compression in sensor networks and video compression where complexity is shifted from the encoder to the decoder. Recently, in [1] we studied Wyner-Ziv (WZ) coding [2, Sec. 14.9] for correlation models where side information may be degraded or absent, and pointed out a relationship with compressed sensing. Such models may be interesting for video coders exploiting WZ concepts [3, 4], where motion-interpolated side information may be degraded or absent due to, *e.g.*, occlusions or new objects appearing. In [1], we presented a theoretical analysis based on (nested) quantization followed by ideal lossless Slepian-Wolf (SW) coding [2, Sec. 14.8], as well as a practical scheme for the erasure case, based on real-valued linear precoding. Although the correlation models in this paper and in [1] are memoryless, which may not correspond to the situation in practical (video) applications, they can be used to study the worst-case performance of more complex models, in which some memoryful process switches between strong and weak correlation.

This paper presents a practical scheme for the degraded side information case with bounded correlation noise, which uses low density parity check (LDPC) codes for the q-ary symmetric channel in the SW stage. The simplest version of the problem considers a continuous memoryless source X (we use X also to denote a source emitting i.i.d. samples of the random variable X), correlated with side information Y available at the decoder only, such that $Y = X$ with probability $1-p_i$, and $Y = X+N_i$ with probability p_i, where N_i is memoryless impulse noise independent of X. In the latter case, we say that the side information is degraded. The decoder does not know whether the side information has been corrupted. This setting can be generalized to include a background noise N_b added to all samples. Correlation models and some theoretic rate-distortion bounds are described in Section 2.

The practical scheme relies on scalar nested quantization followed by SW coding. The q-ary quantization indices U and V of X and Y, respectively, will be modeled as memoryless with $P(V=v|U=u) = 1-p_c$ for $v=u$, and $P(V=v|U=u) = p_c/(q-1)$ for $v \neq u$. The corresponding correlation channel is thus a q-ary symmetric channel with error probability p_c, for which a practical coding approach has been presented in [5]. Section 3 first details the q-ary LDPC approach for syndrome-based SW coding, and then explains its application in a practical two-layer scheme.

Finally, Sections 4 and 5 present first simulation results and an application to compressed sensing, respectively.

This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract n. 216715). The work of C. Weidmann was partly supported by the Austrian Government and by the City of Vienna within the competence center program COMET. The work of F. Bassi and M. Kieffer was partly supported by the French ANR project Essor.

2. CORRELATION MODELS

Consider a Gaussian memoryless source $X \sim \mathcal{N}(0,\sigma_X^2)$, to which the correlation channel adds independent memoryless noise Z to obtain the side information $Y = X+Z$. The *Bernoulli-uniform* (BU) correlation model adds $Z = B \cdot N_i$, a Bernoulli-uniform impulse noise with $P(B=1) = p_i$, $P(B=0) = 1-p_i$ and $N_i \sim \mathcal{U}[0,\beta)$, yielding

$$Y = X + B \cdot N_i \tag{1}$$

as side information available at the decoder. The *uniform Bernoulli-uniform* (UBU) correlation model generalizes this to $Z = N_b + B \cdot N_i$, where $N_b \sim \mathcal{U}[0,\alpha)$, $\alpha < \beta$, is additional uniform background noise, yielding

$$Y = X + N_b + B \cdot N_i. \tag{2}$$

In both models, the Bernoulli source B models the occurrence of impulses degrading the side information. Its realization is known neither at the encoder, nor at the decoder.

Like for the similar correlation models considered in [1], a lower bound on the Wyner-Ziv rate-distortion function (rdf) can be derived assuming that encoder and decoder know the impulse positions and recalling the high-rate property [6]

$$R_{\mathrm{WZ}}(D) \geq R_{\mathrm{X|Y}}(D) \geq h(X|Y) - \tfrac{1}{2}\log_2(2\pi e D).$$

Consider the BU model (1), where the encoder needs to use rate only for the fraction p_i of time when $B=1$. Therefore

$$R_{\mathrm{WZ}}^{\mathrm{BU}}(D) \geq p_i \cdot \big(h(X) + h(N_i) - h(X+N_i) - \tfrac{1}{2}\log_2\left(2\pi e D/p_i\right)\big). \tag{3}$$

For the UBU model (2) one obtains

$$R_{\mathrm{WZ}}^{\mathrm{UBU}}(D) \geq h(X) + (1-p_i)\cdot\big(h(N_i) - h(X+N_i)\big) + p_i \cdot \big(h(N_b+N_i) - h(X+N_b+N_i)\big) - \tfrac{1}{2}\log_2(2\pi e D).$$

An achievable upper bound for the WZ rdf is obtained through high-rate characterization of scalar quantization followed by ideal SW coding. The source symbol is quantized with step ϕ and sent at ideal SW rate; it is decoded error-free using the side information. In the BU case the decoder can detect the presence of impulses by comparing the recovered quantized source with the quantized side information. When no impulse is detected, the decoder reconstructs $\hat{X} = Y$. The resulting overall distortion is smaller than $D = p_i\phi^2/12$ and the asymptotic rate satisfies

$$R_{\mathrm{WZ}}^{\mathrm{BU}}(D) \leq h(p_i) + p_i\left(h(X) + h(N_i) - h(X+N_i) - \tfrac{1}{2}\log_2\left(\tfrac{12D}{p_i}\right)\right), \tag{4}$$

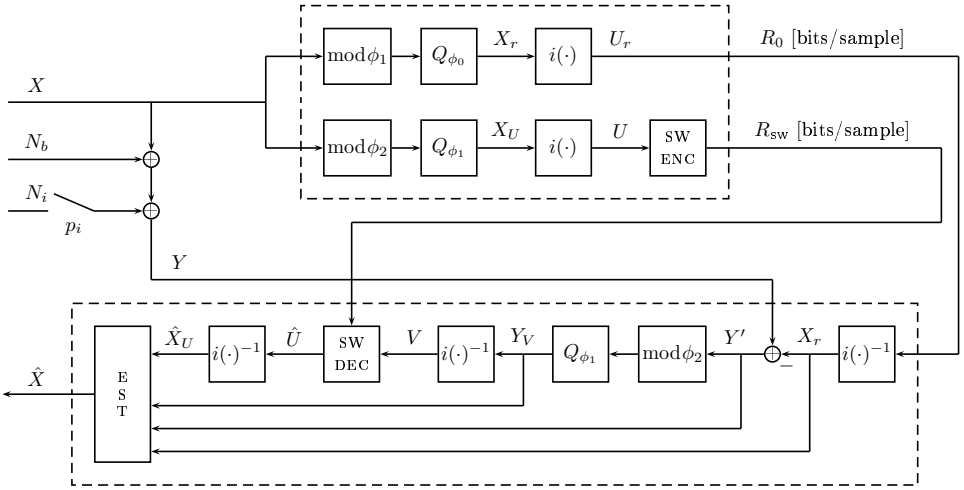

Figure 1: Two-layer encoding scheme.

where $h(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function (all logarithms in this paper are to the base 2). In the UBU case impulses cannot be detected and we obtain

$$R_{\text{WZ}}^{\text{UBU}}(D) \le h(p_i) + h(X) + (1-p_i)\cdot\big(h(N_b) - h(X+N_b)\big) + \\ + p_i\cdot\big(h(N_b+N_i) - h(X+N_b+N_i)\big) - \tfrac{1}{2}\log_2(12D).$$

3. PRACTICAL SCHEME

3.1 Slepian-Wolf coding for q-ary symmetric correlation

As outlined in the introduction, we base our construction on an LDPC SW coding scheme for q-ary discrete memoryless sources (DMS) with a q-ary symmetric correlation channel. Let U be a q-ary DMS with alphabet $\mathbb{Z}_q$, without loss of generality. Assume that the q-ary side information V can be modeled as

$$V = U \oplus W,$$

where $\oplus$ denotes $\bmod q$ addition and W is some memoryless noise with alphabet $\mathbb{Z}_q$. Further assume $P(W=0) = 1-p$, corresponding to the perfect side information case. It is easy to show that uniform distribution of the error value (when the side information is degraded), $P(W=w) = p/(q-1)$ for $w \neq 0$, maximizes the Slepian-Wolf rate, which becomes

$$H(U|V) = H(U \oplus -V|V) = H(W) = h(p) + p\log(q-1). \quad (5)$$

We call this *q-ary symmetric correlation* in reference to the corresponding channel. Its correlating noise W represents a *worst case*, since all error patterns with a given number of errors will be equally likely, *e.g.* for t errors in n positions, all $\binom{n}{t}(q-1)^t$ error patterns will have the same probability $(1-p)^{n-t}p^t(q-1)^{-t}$. Thus it is intuitively clear that a Slepian-Wolf code for this worst-case correlation will work for any noise W, as long as $P(W=0) = 1-p$. In other words, it is an *universal* code for the class of correlating impulse noise with impulse probability p, *i.e.* $P(W \neq 0) = p$. The price for this universality is the redundancy of $h(p) + p\log(q-1) - H(W)$ bits per symbol. Notice that $H(W) = h(p) + pH(W')$, where W' corresponds to W "restricted" to its $q-1$ nonzero values.

In the following we consider only the case $q = 2^m$ with $m \in \mathbb{N}$. Our approach is based on the LDPC coding scheme for the q-ary symmetric channel (q-SC) presented in [5], which we turn into a Slepian-Wolf scheme.

The q-ary channel input and output symbols are represented by binary vectors $\mathbf{u} = [u_1, u_2, \ldots, u_m]^T$ and $\mathbf{v} = [v_1, v_2, \ldots, v_m]^T$, respectively. The q-SC coding scheme in [5] is based on a binary LDPC code with source block size $K = mk$ bits; only the message-passing decoder needs to be modified in order to take into account the underlying q-SC. (For the basics of LDPC codes and message-passing decoding we refer to [7, Ch. 4].)

The modification into a Slepian-Wolf scheme is similar to the syndrome approach in [8]. Given the binary parity check matrix H, the encoder computes the syndrome $\mathbf{s} = \mathrm{H}\underline{\mathbf{u}}(:)$ (the colon $(:)$ denotes the necessary binary serialization of $\underline{\mathbf{u}}$). The message-passing decoder uses $\mathbf{s}$ to change the sign of the log-likelihood ratio check-node messages whenever the corresponding syndrome bit is 1. When the side information $\underline{\mathbf{v}}(:)$ is placed at its channel input, the decoder will thus search the closest $\underline{\hat{\mathbf{u}}}$ having syndrome $\mathbf{s}$.

An important point is that the code is *not* split into m bit planes, the $K = mk$ bits composing the k symbols are all encoded and decoded together, at the price of an $m = \log q$-fold complexity increase in the variable node decoder (see [5] for details). The advantage of this is that no separate optimization of bit planes is needed, and the decoding error decay (waterfall) is much steeper thanks to the longer effective block length. This distinguishes our work from the more general approach used in [9], which requires separate optimization of LDPC codes for each bit plane.

In summary we obtain a universal SW scheme for a pair of correlated q-ary memoryless sources, where the correlation is such that the sources emit equal symbols with probability $1-p$. If p is small, also the penalty (the redundancy due to not uniformly-distributed error values) for this universality will be small. Such a scheme could also have been built using punctured turbo codes [10]; however, that approach may have much higher decoding complexity due to the necessary bit/symbol marginalization (a factor $\mathcal{O}(q)$ instead of $\mathcal{O}(\log q)$) and code optimization is not as straightforward compared to our LDPC approach, for which EXIT charts can be used.

3.2 Two-layer scheme

The main idea in constructing a two-layer scheme is to have a first layer that removes the impulse noise using a Slepian-Wolf code for q-ary symmetric correlation, yielding a first rough estimate of the source, which will then be *refined* by a second layer. A block diagram is shown in Figure 1.

The scheme is based on a doubly nested scalar quantizer, defined by three cell sizes $\phi_0 < \phi_1 < \phi_2$ with integer nesting ratios (i.e., $\phi_1/\phi_0 \in \mathbb{N}$ and $\phi_2/\phi_1 \in \mathbb{N}$). Nested quantization has been proposed for both theoretical analysis and practical implementation of Wyner-Ziv coding (see [9] and references therein). This work extends that approach to doubly nested lattices in order to cope with

the impulse noise.

The quantization operation $Q_\phi(\cdot)$ and the modulo operation $\bmod \phi$ are defined such that $x = Q_\phi(x) + (x \bmod \phi)$ holds for all $x \in \mathbb{R}$. This can be achieved for example by $Q_\phi(x) = \lfloor x/\phi \rfloor \phi$ and $x \bmod \phi = x - Q_\phi(x)$. The fact that all nesting ratios are integers is important, since it allows certain quantization and modulo operations to commute, *e.g.*, $Q_{\phi_1}(x) \bmod \phi_2 = Q_{\phi_1}(x \bmod \phi_2)$.

Layer 1 computes $X_U = Q_{\phi_1}(X \bmod \phi_2)$ and the quantization index $U = X_U/\phi_1$, which has alphabet size $q = |\mathscr{U}| = \phi_2/\phi_1$. U is then sent with rate $R_1 = \log_2 q$ to a q-SC SW encoder for error probability p_c, to be computed later. The SW encoder operates on blocks of n indices U and outputs a (shorter) block of syndrome symbols S, which are sent to the decoder.

Layer 2 computes the refinement $X_r = Q_{\phi_0}(X \bmod \phi_1)$ and sends the quantization index $U_r = X_r/\phi_0$, whose rate is $R_0 = \log_2(\phi_1/\phi_0)$, directly to the decoder.

The decoder starts by computing $Y' = Y - X_r$, which given the UBU correlation model (2) is $Y' = X - X_r + N_b + B \cdot N_i$. If we assume that f_X is constant over a small cell of size ϕ_0, we may decompose the source as $X = Q_{\phi_1}(X) + X_r + N_r$, where $N_r \sim \mathscr{U}[0,\phi_0)$ is uniform quantization noise. Then the decoder quantizes the side information to $Y_V = Q_{\phi_1}(Y' \bmod \phi_2)$ and passes the quantization index $V = Y_V/\phi_1$ to the SW decoder, which uses it together with the syndrome S to produce an estimate $\hat{U}$ of the index U. Notice that the q-SC SW subsystem needs only to know the probability $p_c = \Pr\{V \neq U\}$ to operate correctly. In the following we assume perfect SW decoding, that is $\hat{U} = U$ and thus $\widehat{X}_U = X_U$.

In order for the reconstruction of the coarse nested quantization to be error-free, we need $Q_{\phi_1}(Y') - Q_{\phi_1}(X) < \phi_2$, which guarantees that the noise does not leave the coarsest quantization cell (Layer 1 operates with "units" of size ϕ_1 within a cell of size ϕ_2). We have

$$\begin{aligned} Q_{\phi_1}(Y') - Q_{\phi_1}(X) &= Q_{\phi_1}(X - X_r + N_b + BN_i) - Q_{\phi_1}(X) \\ &= Q_{\phi_1}(Q_{\phi_1}(X) + N_r + N_b + BN_i) - Q_{\phi_1}(X) \\ &= Q_{\phi_1}(X) + Q_{\phi_1}(N_r + N_b + BN_i) - Q_{\phi_1}(X) \end{aligned}$$

and therefore $\phi_2 = \phi_0 + \alpha + \beta$ will be sufficient. From this we immediately obtain

$$\phi_0 = \frac{\alpha + \beta}{2^{R_0+R_1} - 1},$$

as well as $\phi_1 = 2^{R_0}\phi_0 = 2^{-R_1}\phi_2$. Now also the error probability p_c for the q-SC SW stage can be computed:

$$p_c = \begin{cases} 1 - (1-p_i)\frac{2\phi_1 - \phi_0}{2\alpha} - p_i \frac{\phi_0^2/3 + \phi_1^2 - \phi_1\phi_0}{2\alpha\beta}, & \phi_1 < \alpha, \\ p_i + (1-p_i)\frac{(\phi_1 - \alpha - \phi_0)^2}{2\alpha\phi_0} & \\ \quad - p_i\left(\frac{2\phi_1 - \alpha - \phi_0}{2\beta} + \frac{(\alpha + \phi_0 - \phi_1)^3}{6\alpha\beta\phi_0}\right), & \alpha \le \phi_1 < \alpha + \phi_0, \\ p_i\left(1 - \frac{2\phi_1 - \alpha - \phi_0}{2\beta}\right), & \alpha + \phi_0 \le \phi_1 < \beta. \end{cases}$$

Since all encoding operations are carried out $\bmod \phi_2$ (or even $\bmod \phi_1$ for the refinement), the crucial step in the reconstruction of the source is the estimation of $Q_{\phi_2}(X)$, which is possible thanks to the above choice of ϕ_2. Consider

$$\begin{aligned} (Y_V - X_U) \bmod \phi_2 &= \big[Q_{\phi_1}((X - X_r + N_b + BN_i) \bmod \phi_2) \\ &\quad - Q_{\phi_1}(X \bmod \phi_2)\big] \bmod \phi_2 \\ &= Q_{\phi_1}\big((Q_{\phi_1}(X) + N_r + N_b + BN_i) \bmod \phi_2 \\ &\quad - Q_{\phi_1}(X \bmod \phi_2)\big) \bmod \phi_2 \\ &= Q_{\phi_1}(N_r + N_b + BN_i), \end{aligned}$$

where the last equality follows again thanks to the choice of ϕ_2, and

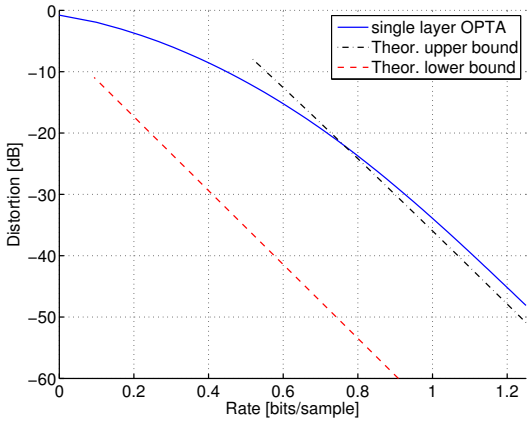

Figure 2: Asymptotic bounds and single layer OPTA for BU model with $p_i = 0.1$, $\beta = 10$, $\sigma_X^2 = \beta^2/12$.

insert this into

$$\begin{aligned} &Y' - (Y_V - X_U) \bmod \phi_2 \\ &\quad = Y - X_r - Q_{\phi_1}(N_r + N_b + BN_i) \bmod \phi_2 \\ &\quad = Q_{\phi_1}(X) + N_r + N_b + BN_i - Q_{\phi_1}(N_r + N_b + BN_i) \\ &\quad = Q_{\phi_1}(X) + (N_r + N_b + BN_i) \bmod \phi_1. \end{aligned}$$

The last expression is smaller than $Q_{\phi_1}(X) + \phi_1$ and therefore allows to compute $Q_{\phi_2}(X)$, yielding the reconstruction formula

$$\widehat{X} = Q_{\phi_2}(Y' - (Y_V - \widehat{X}_U) \bmod \phi_2) + \widehat{X}_U + X_r.$$

Using the ideal SW rate R_{SW} (5), the total rate becomes

$$R(R_0, R_1) = R_0 + h(p_c) + p_c \log_2(2^{R_1} - 1), \tag{6}$$

while the distortion is

$$D(R_0, R_1) = \frac{\phi_0^2}{12} = \frac{(\alpha+\beta)^2}{12}(2^{R_0+R_1} - 1)^{-2}. \tag{7}$$

under the above assumption on f_X (if the standard deviation of X is not much larger than ϕ_0, a subtractive dither might be introduced).

When $R_0 = 0$, this scheme reduces to a single layer, which is all that is needed in the BU case (1), where $\alpha = 0$. Then the "channel" probability $p_c = p_i(1 - \phi_1/\beta)$ and the distortion $D(R_1) = p_i\phi_1^2/12$, since only the samples with degraded side information will be noisy. Figure 2 shows the optimal performance theoretically attainable (OPTA) curve $(R(R_1), D(R_1))$ together with theoretical upper (4) and lower (3) bounds. Notice that the upper and lower bounds are asymptotic (low D) bounds on the rate-distortion function, while the OPTA measures the ideal performance of the two-layer system in Figure 1. The rate gap between high-rate OPTA and (4) is thus the result of an inherent suboptimality of the two-layer design; it equals $p_i(h(X + N_i) - h(X))$.

Some improvements at low rate could be obtained by modifying the scheme such that the SW decoder operates directly with Y, instead of its quantized version. However, quantizing Y has the advantage of making this scheme easily generalizable beyond the asymmetric WZ setup. In some applications, it might be more desirable to compress both X and Y independently, but still be able to exploit the (sparse) impulsive nature of the correlating noise Z. In the generalized setting, Y is encoded in the same fashion as X (see Figure 1), but possibly with different SW rate $R_{1,Y}$. To achieve arbitrary integer SW rates $R_{1,X}$ and $R_{1,Y}$, it then suffices to modify the SW subsystem along the lines of [11] (the refinement rate R_0 will be the same for both X and Y).

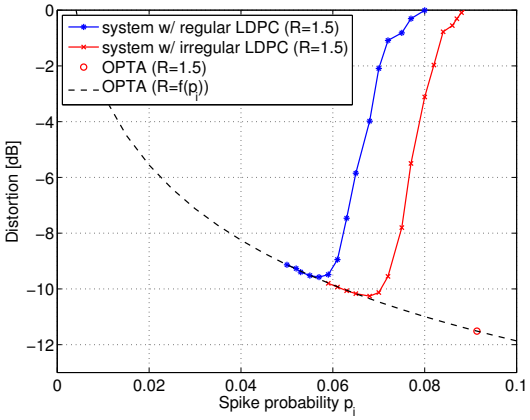

Figure 3: Performance of scheme based on rate $3/4$ LDPC ($R_0 = 1$ b/s, $R_1 = 2$ b/s; total rate $R = R_0 + (1-3/4)R_1 = 1.5$ b/s).

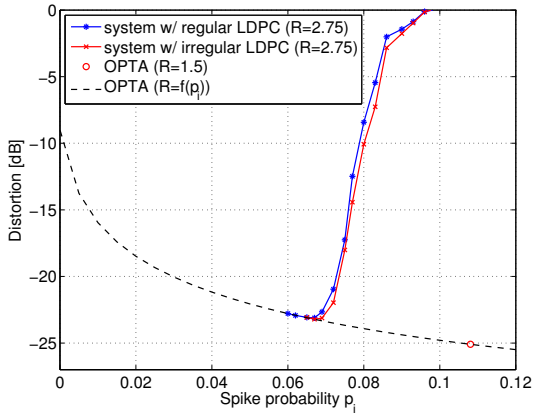

Figure 4: Performance of scheme based on rate $3/4$ LDPC ($R_0 = 2$ b/s, $R_1 = 3$ b/s; total rate $R = R_0 + (1-3/4)R_1 = 2.75$ b/s).

To apply this two-layer approach to unbounded correlation noise, such as the Gaussian-Bernoulli-Gaussian (GBG) model introduced in [1], one needs to consider the additional distortion caused by the (necessarily) nonzero probability of the noise Z leaving the cells of size ϕ_2. This can be counteracted by increasing ϕ_2, which however decreases the Wyner-Ziv coding efficiency and still results in a distortion floor for large refinement rate R_0 [9]. To overcome that problem, one may increase the dimension (so ϕ_2 becomes a lattice) and/or use an additional layer of Slepian-Wolf coding to fix the (rare) cases in which Z overshoots, as introduced in [9].

4. SIMULATION RESULTS

Simulations were performed using an unoptimized regular $(3,12)$ binary LDPC code of length $N = 12000$ bits, which has channel coding rate $3/4$ and thus syndrome rate $1/4$. A second set of simulations was made using AWGN-optimized irregular LDPC codes of the same length and rate, with maximum left degree 15, taken from [12]. A standard binary message-passing decoder was modified into a q-SC Slepian-Wolf decoder as outlined in Sec. 3.1 and then embedded into a two-layer scheme (Sec. 3.2). The source was chosen as $X \sim \mathcal{N}(0, 50^2)$, and the side information was UBU-correlated with $\alpha = 1$, $\beta = 10$ and variable impulse probability p_i. The latter is necessary to show the convergence behavior of the LDPC decoder; however, it makes comparisons more difficult, as will be explained. Figure 3 shows relative distortion vs. impulse probability for a system with $R_0 = 1$, $R_1 = 2$ and thus total rate $R = R_0 + (1-3/4)R_1 = 1.5$ bits per sample (b/s). Distortion is shown relative to the linear MMSE achievable at rate $R = 0$, which in this Wyner-Ziv setting is also known as conditional variance,

$$D_0 = \sigma^2_{X|Y} = \frac{\sigma^2_X \sigma^2_Z}{\sigma^2_X + \sigma^2_Z},$$

where $\sigma^2_Z = \frac{\alpha^2}{12} + p_i \beta^2 \left(\frac{1}{3} - \frac{p_i}{4}\right)$. This dependency of D_0 on p_i makes comparisons somewhat problematic, since choosing R_0 and R_1 fixes the quantizer cell sizes ϕ_0, ϕ_1, ϕ_2 and thus the distortion $D = \phi_0^2/12$ as well. This means that decreasing p_i will increase the relative distortion, as can be seen from the dashed curve in Figure 3, which represents the OPTA curve for a system with fixed scalar quantizers (determined by R_0, R_1) and an ideal Slepian-Wolf subsystem. Total rate R varies along the OPTA curve, the point with $R = 1.5$ b/s has been marked with a circle. At the point where the regular LDPC decoder converges ($p_i = 0.056$), the OPTA system has rate 1.34 b/s. For the irregular code, these are $p_i = 0.066$ and 1.39 b/s, respectively.

Figure 4 shows the same quantities for a system with $R_0 = 2$, $R_1 = 3$, and total rate $R = R_0 + (1-3/4)R_1 = 2.75$ b/s. At the point where the regular LDPC decoder converges ($p_i = 0.066$), the OPTA system has rate 2.50 b/s. In this configuration, the irregular code has only a marginal advantage, since for symbols of $m = R_1 = 3$ bits, the variable-node decoder (EXIT) characteristic is already very different from that of a binary channel input alphabet.

5. APPLICATION TO COMPRESSED SENSING

In [1], we suggested that a theoretical WZ scheme for impulse-noise degraded side information could be used for compressed sensing (CS) [13, 14]. Here we outline how the above practical scheme can be applied and discuss the differences with common CS techniques, as well as standard compression of sparse sources.

A bare bones example of a CS problem is to compress a signal vector $\mathbf{x} \in \mathbb{R}^N$ of the form $\mathbf{x} = \Psi \mathbf{s}$, where Ψ is an orthonormal N-by-N matrix and $\mathbf{s} \in \mathbb{R}^N$ has at most K non-zero components (we say that $\mathbf{x}$ is K-sparse with respect to Ψ). The problem is to determine a compression mechanism for $\mathbf{x}$ without using the sparsifying basis Ψ at the encoder (for complexity reasons) and to characterize its rate-distortion behavior for an appropriate random model of $\mathbf{s}$. A distributed CS (DCS) problem might consider a signal which is known to have a sparse difference with respect to a reference signal $\mathbf{y}$ (side information) available only at the decoder.

In the simple case $\Psi = I_N$, when $\mathbf{x} = \mathbf{s}$ is sparse in the standard basis, and the components of $\mathbf{x}$ are bounded, there is an immediate connection with the UBU model: consider $\mathbf{x}$ as a vector of i.i.d. samples of UBU noise with $p_i = K/N$ and BN_i the nonnegative (without loss of generality) sparse signal components, while N_b is some background noise (this actually generalizes the concept of K-sparsity). Then a CS scheme can be built using the two-layer scheme (of course one could use a standard entropy-coded scalar quantizer in this non-distributed setting—the qualitative differences of these two approaches will be explained below). The decoder uses the trivial side information $\mathbf{y} = \mathbf{0}$ to decode the quantized values $\hat{\mathbf{x}}$ (here the roles of source and correlation noise are reversed), while in the more interesting case of DCS, it will simply use the provided $\mathbf{y}$. Notice that the system is basically a fixed-rate (nested) quantizer followed by a syndrome (Slepian-Wolf) source code, where the qualifiers in parentheses apply in the DCS setting. In the following we refer to this as quantized syndrome coding (QSC). A key observation is that the rate (6) is essentially linear in the *coding sparsity* p_c, which can be made equal to *source sparsity* p_i by choice of R_1 (also, $p_c \leq p_i$ in the BU case).

It is possible to adapt the non-distributed CS scheme to more general source distributions using a companding quantizer [15]. For example, a strictly sparse Bernoulli-Gaussian (BG) source with pa-

rameters p_i and σ_i^2 can be encoded with distortion [16, p. 2329]

$$D(R_1) \cong \frac{p_i}{12} 6\pi\sqrt{3}\sigma_i^2 2^{-2R_1}$$

and, from (5), rate $R(R_1) \cong h(p_i) + p_i R_1$, if the quantizer rate R_1 is not too small. This rate expression matches the "position plus value" code used in quantized nonlinear approximation (NLA) [17], where for a length N block, $Nh(K/N)$ bits encode the positions of $K = Np_i$ nonzero samples, of which each is quantized with R_1 bits.

Since the performance of this QSC system is equivalent to that of standard scalar entropy-coded quantized NLA of sparse sources, a comparison is in order. The standard system has generally higher encoder complexity and needs buffering to smooth the variable transmission rate (both due to entropy coding (EC)), it has low delay (depending on EC and buffering) and bounded distortion per sample. The QSC system has lower encoder complexity (per sample) and fixed transmission rate (needing no buffers), but larger delay (due to block size) and nonzero block error probability, which affects the distortion on a block basis.

Comparing with common CS methods based on random projections and linear programming reconstruction [13, 14] is also insightful. CS is generally formulated as a sampling problem, where the key parameter is the sampling ratio needed to reconstruct a signal with a given sparsity. Here we consider the same as an information-theoretic compression problem, where the key figure is the number of bits per sample needed to achieve a certain reconstruction distortion. Notice that in the sampling setting, distortion vanishes if the reconstruction is perfect. Since a real value may need an infinite number of bits to be encoded, it is only by allowing some distortion that a compression result can be obtained. In the limit of low distortion (high rate), the "sampling ratio" R/R_1 (corresponding to the ratio of real-valued samples) tends to $p_i = K/N$, which is also the best possible in the CS setting. When the trivial sparsifying basis $\Psi = I_N$ is assumed, which is often the case also in the CS literature, these two approaches to sparse signal compression share striking similarities. In the CS setting, $\mathbf{x}$ is projected onto a random subspace using a fat matrix Φ. Then, if digital transmission is used between source encoder (sensor) and decoder, $\Phi\mathbf{x}$ needs to be quantized and reconstructed, before $\mathbf{x}$ can be recovered using, *e.g.*, linear programming techniques. In the LDPC-based QSC setting, $\mathbf{x}$ is quantized first, then its digital representation is projected using a fat sparse parity check matrix H. The LDPC decoder recovers this representation by solving a system of binary linear equations and finally the quantized $\mathbf{x}$ is recovered. Assuming scalar quantization, the CS approach needs to deal with real-valued (analog) matrix multiplication, while the LDPC-QSC approach uses binary (digital) matrix multiplication.

In the more general case, when $\mathbf{x}$ is sparse in a non-trivial basis $\Psi \neq I_N$, the QSC approach could only used if the overall codebook (of size 2^R) were invariant under the rotation Ψ, which allows to encode without knowing Ψ. It needs to be investigated whether this condition is approximately satisfied if the rate is high enough, such that the rotation Ψ can be "undone" in syndrome space.

A view on compressed sensing similar to ours was recently presented in [18], which however considers codes over the reals (similar in spirit to the precoding in [1], which draws from [19]), while here we propose digital codes with low-complexity decoding algorithms. Finally, we remark that the overview article [20] presents a different and more general view of quantization plus Slepian-Wolf methods for compressed sensing.

6. CONCLUSION

This paper introduced a practical scheme for Wyner-Ziv coding with side information degraded by bounded impulse noise. The key ingredient of this scheme is an LDPC-based Slepian-Wolf code for the q-ary symmetric correlation channel, which corrects discrete impulse errors (with any distribution) up to a chosen frequency p_c. The main advantage of this method is that in the high-rate limit, it needs only half as much syndrome information as a comparable real-valued error correction approach (as, *e.g.*, in [1], which uses techniques from [19]); the reason behind this being that the q-ary symmetric channel capacity tends to $1 - p_c$ for large alphabet sizes q. Therefore this new scheme is also interesting for compressed sensing applications, for which the equivalent critical sampling rate $p_c = K/N$ can be approached arbitrarily closely.

Acknowledgments

The authors would like to thank Gottfried Lechner for kindly providing the LDPC toolbox used in the simulations.

References

[1] F. Bassi, M. Kieffer, and C. Weidmann, "Source coding with intermittent and degraded side information at the decoder," in *Proc. ICASSP*, Las Vegas, NV, USA, Mar. 30 – Apr. 4, 2008.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

[3] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[4] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Trans. Image Proc.*, vol. 16, no. 10, pp. 2436–2448, Oct. 2007.

[5] C. Weidmann, "Coding for the q-ary symmetric channel with moderate q," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, Canada, Jul. 2008.

[6] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2073–2084, Nov. 1996.

[7] R. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, to appear 2008, see also http://lthcwww.epfl.ch/mct/index.php.

[8] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002.

[9] 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. Inform. Theory*, vol. 52, no. 10, pp. 4358–4379, Oct. 2006.

[10] Y. Zhao and J. Garcia-Frias, "Data compression of correlated non-binary sources using punctured turbo codes," in *Proc. Data Compression Conference (DCC)*, Apr. 2002.

[11] N. Gehrig and P. L. Dragotti, "Symmetric and asymmetric Slepian-Wolf codes with systematic and nonsystematic linear codes," *IEEE Commun. Lett.*, vol. 9, pp. 61–63, Jan. 2005.

[12] LdpcOpt, http://lthcwww.epfl.ch/research/ldpcopt/.

[13] E. Candès and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[14] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[15] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, Jul. 1948.

[16] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[17] C. Weidmann and M. Vetterli, "Rate distortion behavior of threshold-based nonlinear approximations," in *Proc. Data Compression Conference (DCC)*, Mar. 2000, pp. 333–342.

[18] F. Zhang and H. D. Pfister, "Compressed sensing and linear codes over real numbers," in *Proc. 2008 Workshop on Inform. Theory and Appl. (ITA@UCSD)*, La Jolla, CA, USA, Feb. 2008.

[19] A. Gabay, M. Kieffer, and P. Duhamel, "Joint source-channel coding using real BCH codes for robust image transmission," *IEEE Trans. Image Proc.*, vol. 16, pp. 1568–1583, Jun. 2007.

[20] V. K Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Processing Mag.*, vol. 25, no. 2, Mar. 2008.