Quantization of Memoryless Sources Over Discrete Markov Channels†

Nam Phamdo[1], *Fady Alajaji*[2], *Nariman Farvardin*[2] *and Thomas E. Fuja*[2]

[1]Department of Electrical Engineering
State University of New York at Stony Brook, Stony Brook, NY 11794-2350

[2]Department of Electrical Engineering and Institute for Systems Research
University of Maryland, College Park, MD 20742

ABSTRACT

Joint source-channel coding for real-valued memoryless sources and binary Markov channels is considered. The channel is an additive-noise channel where the noise process is an M-th order Markov chain. We examine the special case where the noise sample Z_i depends only on the *sum* of the previous M noise samples. Two joint source-channel coding schemes are considered. The first is a channel-optimized vector quantizer – optimized for both source and channel. The second scheme consists of a simple scalar quantizer and a maximum a posteriori (MAP) detector. In this scheme, it is assumed that the scalar quantizer output has residual redundancy that can be exploited by the MAP detector to combat the correlated channel noise. These two schemes are then compared against two schemes which use interleaving. Numerical results show that the proposed schemes outperform the interleaving schemes. In some instances, the gain is more than 5 dB.

I. Introduction

Source and channel coding are two problems that have traditionally been dealt with independently. This is due mainly to Shannon's source-channel separation principle [1], [2], which states that the two problems can be treated separately without loss of optimality. However, the separation principle holds only in the asymptotic case — when delay and complexity are not constrained. Recent works [3], [4], [5] have shown that, when delay and/or complexity are constrained, treating these problems jointly (i.e., joint source-channel coding) may result in improved performance over the traditional technique of tandem source-channel coding.

Most of the previous work on joint source-channel coding have assumed that the channel is memoryless, disregarding the fact that real-world communication channels often have memory. In this work, we will consider two joint source-channel coding schemes for memoryless sources and for channels with memory. More specifically, the source is assumed to be a real-valued, independent and identically distributed (i.i.d.) sequence of random variables and the channel is assumed to be a binary stationary ergodic M-th order Markov channel derived from the Polya contagion urn model [6]. This is an additive-noise channel where the noise sample, Z_i, depends only on the *sum* of the previous M noise samples $(Z_{i-1}, Z_{i-2}, \ldots, Z_{i-M})$.

We first consider the design of a k-dimensional, rate R bits/sample channel-optimized vector quantizer (COVQ) [7], [8] for the given source and channel. The COVQ encoder output is transmitted over the Markov channel. For each block of k source samples, the COVQ encoder produces kR bits for transmission. We assume that kR is large enough with respect to M so that the memory in the channel can have an effect in kR channel uses. Thus, by a proper design of the COVQ, we exploit the intra-block memory of the channel — but not the inter-block memory. The COVQ design algorithm is a straightforward extension of the algorithm described in [7] and [8], where the $2^{kR} \times 2^{kR}$ channel transition matrix is now given in terms of the transition probabilities of the Markov channel.

We then exploit both intra-block and inter-block memories of the channel. Here, we consider a scalar quantizer (SQ) designed for the noiseless channel. The SQ output distribution is assumed to be non-uniform so that its entropy (in bits/channel use) is strictly less than the channel capacity (bits/channel use). After a proper assignment of binary indices to the SQ output, we transmit the indices directly over the channel. At the receiver, we exploit the non-uniformity of the SQ output and the memory of the channel through the use of a sequence maximum a posteriori (MAP) detector. The output of the MAP detector is then fed to the SQ decoder. This is analogous to previous works on MAP detection of a Markov source over a memoryless channel [9], [10].

The performances of the two proposed schemes are compared against the performances of two interleaving schemes. In the interleaving systems, the Markov channel is rendered memoryless by an interleaver and de-interleaver[3]. Here, we assume that the source and channel codes are designed for the memoryless channel. Thus, the purpose of the interleaver and de-interleaver is to convert the Markov channel (with memory) into a memoryless channel. In the first interleaving scheme, we consider

†The work of the last three authors is supported in part by National Science Foundation grant NSFD CD-8803012.

[3]* It is assumed that the interleaver and de-interleaver are *ideal* so that the Markov channel is perfectly rendered memoryless.

a COVQ designed for a memoryless channel with the same bit error rate as the Markov channel. This COVQ is then used over the interleaved channel (combination of interleaver, Markov channel and de-interleaver). This system is compared against the COVQ designed for the Markov channel. In the second interleaving system, we consider an SQ with its output transmitted over a memoryless (interleaved) channel. We call this system "SQ-Interleaved". This is compared against "SQ-MAP" where the MAP detector is designed for the Markov channel.

The rest of this paper is organized as follows. In Section II, we present the Markov channel model. The two joint source-channel coding schemes are described in Section III. Simulation results for i.i.d. generalized Gaussian sources are provided in Section IV. In Section V, comparisons between the proposed schemes and the corresponding interleaving schemes are made. Finally, the conclusions are given in Section VI.

II. Channel Model

Consider a discrete channel with memory, with common input, noise and output binary alphabet and described by the following equation: $Y_i = X_i \oplus Z_i$, for $i = 1, 2, 3, \ldots$ where:

- $\oplus$ represents the addition operation modulo 2.
- The random variables X_i, Z_i and Y_i are respectively the input, noise and output of the channel.
- $\{X_i\} \perp \{Z_i\}$, i.e., the input and noise sequences are independent from each other.
- The noise process $\{Z_i\}_{i=1}^{\infty}$ is a homogeneous stationary mixing (hence ergodic) Markov process of order M. By this we mean that the noise sample, Z_i, depends only on the previous M noise samples, i.e.,

$$\begin{aligned}\Pr\{Z_i = e_i | Z_1 = e_1, \ldots, Z_{i-1} = e_{i-1}\} = \\ \Pr\{Z_i = e_i | Z_{i-M} = e_{i-M}, \ldots, Z_{i-1} = e_{i-1}\}.\end{aligned}$$

We assume that the marginal distribution of the noise process is given by

$$\Pr\{Z_i = 1\} = \epsilon = 1 - \Pr\{Z_i = 0\},$$

where $\epsilon \in [0, 1/2)$ is the channel bit error rate (BER). Furthermore, we assume that the process $\{Z_i\}$ is generated by the finite-memory contagion urn model described in [6]. According to this model, the noise sample Z_i depends only on the *sum* of the previous M noise samples. Thus, for $i \geq M+1$,

$$\begin{aligned}\Pr\{Z_i = 1 | Z_{i-M} = e_{i-M}, \ldots, Z_{i-1} = e_{i-1}\} \\ = \Pr\{Z_i = 1 | \sum_{j=i-M}^{i-1} Z_j = \sum_{j=i-M}^{i-1} e_j\} \\ = \frac{\epsilon + (\sum_{j=i-M}^{i-1} e_j)\delta}{1 + M\delta},\end{aligned}$$

where $e_j = 0$ or 1, for $j = i-M, \ldots, i-1$. The positive parameter δ determines the amount of correlation in $\{Z_i\}$. The correlation coefficient of the noise process is $\delta/(1+\delta)$. Note that if $\delta = 0$, the noise process $\{Z_i\}$ becomes i.i.d. and the resulting additive noise channel becomes a binary symmetric channel (BSC).

A. Distribution of the Noise

For an input block $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and an output block $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, we denote the block channel transition probability matrix $\Pr\{\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}\}$ by $Q(\mathbf{y}|\mathbf{x})$.

- For block length $n \leq M$, we have [6]:

$$Q(\mathbf{y}|\mathbf{x}) = L(n, d, \epsilon, \delta),$$

where

$$L(n, d, \epsilon, \delta) = \frac{\left[\prod_{i=0}^{d-1}(\epsilon + i\delta)\right]\left[\prod_{j=0}^{n-d-1}(1 - \epsilon + j\delta)\right]}{\left[\prod_{l=0}^{n-1}(1 + l\delta)\right]},$$

and $d = d_H(\mathbf{x}, \mathbf{y})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$.

- For $n \geq M+1$, we obtain [6]:

$$\begin{aligned}Q(\mathbf{y}|\mathbf{x}) &= \Pr\{\mathbf{Z} = \mathbf{e}\} \\ &= L(M, s, \epsilon, \delta) \prod_{i=M+1}^{n} \left[\frac{\epsilon + s_i\delta}{1 + M\delta}\right]^{e_i} \left[1 - \frac{\epsilon + s_i\delta}{1 + M\delta}\right]^{1-e_i} \quad (1)\end{aligned}$$

where $\mathbf{e} = (e_1, e_2, \ldots, e_n)$, $e_i = x_i \oplus y_i$, $s = e_1 + \cdots + e_M$ and $s_i = e_{i-1} + \cdots + e_{i-M}$.

Note that the channel is entirely described by ϵ, δ and M.

B. Capacity of the Channel

The capacity C of this channel is given by [6]:

$$C = 1 - \sum_{s=0}^{M} \binom{M}{s} L(M, s, \epsilon, \delta) h_b\left(\frac{\epsilon + s\delta}{1 + M\delta}\right)$$

where $h_b(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function. Note that C is monotonically increasing with δ (for fixed ϵ, M) and M (for fixed ϵ, δ). It is monotonically decreasing with ϵ (for fixed δ, M).

III. Joint Source-Channel Coding Schemes

A. COVQ and COSQ

The ensuing formulation of COVQ follows that of [8]. Consider a real-valued i.i.d. source, $\mathcal{V} = \{V_i\}_{i=1}^{\infty}$, with probability density function (p.d.f.) $f(v)$. The source is to be encoded by a k-dimensional, n-bit COVQ whose output is to be transmitted over the binary Markov channel. The encoding system, depicted in Figure 1, consists of an encoder mapping, γ, and a decoder mapping, β. The encoder mapping $\gamma : \mathbb{R}^k \mapsto \{0,1\}^n$ is described in terms of a partition $\mathcal{P} = \{S_\mathbf{x} \subset \mathbb{R}^k : \mathbf{x} \in \{0,1\}^n\}$ of $\mathbb{R}^k$ according to

$$\gamma(\mathbf{v}) = \mathbf{x} \quad \text{if } \mathbf{v} \in S_\mathbf{x}, \;\; \mathbf{x} \in \{0,1\}^n,$$

Figure 1. Block Diagram of a COVQ System.

where $\mathbf{v} = (v_1, v_2, \ldots, v_k)$ is a block of k successive source samples. The channel takes an input sequence $\mathbf{x}$ and produces and output sequence $\mathbf{y}$. It is given in terms of the block channel transition matrix $Q(\mathbf{y}|\mathbf{x})$. Finally, the decoder mapping $\beta : \{0,1\}^n \mapsto \mathrm{I\!R}^k$ is described in terms of a codebook $\mathcal{C} = \{\mathbf{c_y} \in \mathrm{I\!R}^k : \mathbf{y} \in \{0,1\}^n\}$ according to

$$\beta(\mathbf{y}) = \mathbf{c_y}, \qquad \mathbf{y} \in \{0,1\}^n.$$

The encoding rate of the above system is $R = n/k$ bits/sample and its average squared-error distortion per sample is given by [8]:

$$D = \frac{1}{k} \sum_{\mathbf{x}} \int_{S_{\mathbf{x}}} f(\mathbf{v}) \left\{ \sum_{\mathbf{y}} Q(\mathbf{y}|\mathbf{x}) ||\mathbf{v} - \mathbf{c_y}||^2 \right\} d\mathbf{v}, \qquad (2)$$

where $f(\mathbf{v}) = \prod_{i=1}^k f(v_i)$ is the k-dimensional source p.d.f. For a given source, channel, k and n, we wish to minimize D by proper choice of $\mathcal{P}$ and $\mathcal{C}$.

From (2), we see that for a fixed $\mathcal{C}$ the optimal partition $\mathcal{P}^* = \{S^*_{\mathbf{x}}\}$ is given by [8]:

$$S^*_{\mathbf{x}} = \left\{ \mathbf{v} : \sum_{\mathbf{y}} Q(\mathbf{y}|\mathbf{x}) ||\mathbf{v} - \mathbf{c_y}||^2 \right.$$
$$\left. \leq \sum_{\mathbf{y}} Q(\mathbf{y}|\tilde{\mathbf{x}}) ||\mathbf{v} - \mathbf{c_y}||^2, \forall \tilde{\mathbf{x}} \in \{0,1\}^n \right\}, \qquad (3)$$

$\mathbf{x} \in \{0,1\}^n$. Similarly, the optimal codebook $\mathcal{C}^* = \{\mathbf{c}^*_{\mathbf{y}}\}$ for a given partition is [8]:

$$\mathbf{c}^*_{\mathbf{y}} = \frac{\sum_{\mathbf{x}} Q(\mathbf{y}|\mathbf{x}) \int_{S_{\mathbf{x}}} \mathbf{v} f(\mathbf{v}) d\mathbf{v}}{\sum_{\mathbf{x}} Q(\mathbf{y}|\mathbf{x}) \int_{S_{\mathbf{x}}} f(\mathbf{v}) d\mathbf{v}}. \qquad (4)$$

The COVQ design algorithm is a straightforward extension of the iterative algorithm in [11]. The algorithm starts out with an initial codebook, $\mathcal{C}^{(0)}$. With this fixed, it finds the optimal partition, $\mathcal{P}^{(1)}$, using (3). With $\mathcal{P}^{(1)}$ fixed, it uses (4) to find the optimal codebook, $\mathcal{C}^{(1)}$. This procedure is repeated until the relative change in distortion is sufficiently small. Note that the average distortion, D, decreases monotonically at each step. Thus, the algorithm is guaranteed to converge to a locally optimal solution (since $D \geq 0$). For $k = 1$, the above system is referred to as channel-optimized scalar quantizer (COSQ). We will assume that $n \geq M + 1$. Therefore, the block channel transition matrix, $Q(\mathbf{y}|\mathbf{x})$, will always be given by (1).

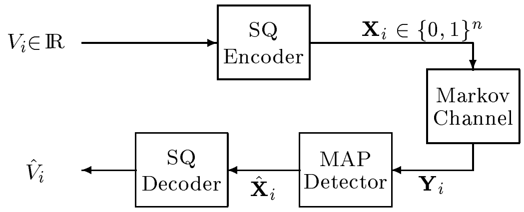

Figure 2. Block Diagram of Joint Source-Channel Coding System Using MAP Detection.

B. MAP Detection

Next consider the system depicted in Figure 2. Here, instead of using COVQ we use a scalar quantizer (SQ). The SQ is also described by γ and β as above — except that $k = 1$ and $R = n$ bits/sample. Instead of optimizing the SQ for the Markov channel, we make use of the residual redundancy of the SQ to combat channel noise. This is in the spirit of the approaches in [5], [9] and [10].

The SQ in Figure 2 is designed using the Lloyd-Max formulation [12], [13] which assumes the channel is noise-free. Since the source, $\mathcal{V} = \{V_i\}_{i=1}^\infty$, is i.i.d., the SQ encoder output, $\{\mathbf{X}_i\}_{i=1}^\infty$, is also i.i.d. Therefore, there is no memory at the encoder output. However, there may be redundancy in the form of a non-uniform distribution on $\mathbf{X}$. Let $p(\mathbf{x}) = \Pr\{\mathbf{X} = \mathbf{x}\}$, $\mathbf{x} \in \{0,1\}^n$ be the encoder output distribution and $H(\mathbf{X}) = -\sum p(\mathbf{x}) \log p(\mathbf{x})$ be the entropy of this distribution. Let $\rho_D = n - H(\mathbf{X})$ be the redundancy (due to the non-uniform distribution) of $\mathbf{X}$. We will assume that $\rho_D > 0$, i.e., $p(\mathbf{x})$ is non-uniform. For a Lloyd-Max scalar quantizer, this is often the case when the source distribution, $f(v)$, is non-uniform (assuming that $R > 1$ bit/sample).

The SQ encoder output is transmitted directly over the channel. At the receiver, a sequence MAP detector is used to exploit the redundancy of $\mathbf{X}$ and the memory of the noise. The MAP detector output is then fed to the SQ decoder. The sequence MAP detector observes a sequence $\mathbf{y}^N = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N) \in \{0,1\}^{nN}$ and makes an estimate of the sequence $\mathbf{x}^N = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \in \{0,1\}^{nN}$ according to

$$\hat{\mathbf{x}}^N = \arg\max_{\mathbf{x}^N} \Pr\{\mathbf{X}^N = \mathbf{x}^N | \mathbf{Y}^N = \mathbf{y}^N\}.$$

It can be easily shown that if $n \geq M$,

$$\hat{\mathbf{x}}^N = \arg\max_{\mathbf{x}^N} \left\{ \log\left[Q(\mathbf{y}_1|\mathbf{x}_1) p(\mathbf{x}_1)\right] \right.$$
$$\left. + \sum_{i=2}^N \log\left[\tilde{Q}(\mathbf{e}_i|\mathbf{e}_{i-1}) p(\mathbf{x}_i)\right] \right\}, \qquad (5)$$

where $\mathbf{e}_i = \mathbf{x}_i \oplus \mathbf{y}_i \in \{0,1\}^n$, $i = 1, 2, \ldots, N$, and $\tilde{Q}(\mathbf{e}_i|\mathbf{e}_{i-1}) = \Pr\{\mathbf{Z}_i = \mathbf{e}_i | \mathbf{Z}_{i-1} = \mathbf{e}_{i-1}\}$. Note that for

$i \geq 2$,

$$\tilde{Q}(\mathbf{e}_i|\mathbf{e}_{i-1}) = \prod_{j=n(i-1)+1}^{ni} \left[\frac{\epsilon + s_j\delta}{1+M\delta}\right]^{e_j} \left[1 - \frac{\epsilon + s_j\delta}{1+M\delta}\right]^{1-e_j},$$

where $\mathbf{e}_{i-1} = (e_{n(i-2)+1}, e_{n(i-2)+2}, \ldots, e_{n(i-1)})$, $\mathbf{e}_i = (e_{n(i-1)+1}, e_{n(i-1)+2}, \ldots, e_{ni})$ and $s_j = e_{j-1} + \cdots + e_{j-M}$.

As expressed in (5), the sequence MAP detector can be implemented using the Viterbi algorithm, where $\mathbf{x}_i$ is the state at time instant i. The trellis has 2^n states with 2^n branches leaving and entering each state. For a branch leaving state $\mathbf{x}_{i-1}$ and entering state $\mathbf{x}_i$, the path metric is $\log[\tilde{Q}(\mathbf{x}_i \oplus \mathbf{y}_i|\mathbf{x}_{i-1} \oplus \mathbf{y}_{i-1})p(\mathbf{x}_i)]$. From here on, this scheme will be referred to as SQ-MAP. We note that the complexity and delay of SQ-MAP is due mainly to the MAP detector.

In some special instances, the output of the MAP detector is identical to its input. In such cases, we say that the MAP detector is useless. As an example, when $M = n = 1$, it is shown in [14] that the MAP detector is useless if

$$\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right] \left[\frac{1-p}{p}\right] \geq 1, \tag{6}$$

where $p = \Pr\{X = 0\} \in (1/2, 1]$. If (6) does not hold, then the sequence MAP detector will be useful for sufficiently large N [14].

In this paper, we are mainly interested in cases where $M = 1$ and $n > 1$. In these cases, little is known about the usefulness of the MAP detector. However, an important factor contributing to the performance of the MAP detector is how the binary codewords are assigned to the SQ quantization levels. This issue will be discussed in the following section.

IV. Numerical Results

In the following, we will assume that the source distribution is given by

$$f(v) = \frac{\alpha\eta(\alpha,\sigma)}{2\Gamma(1/\alpha)} \exp\{-[\eta(\alpha,\sigma)|v|]^{\alpha}\}, \tag{7}$$

where $\eta(\alpha,\sigma) = \sigma^{-1}[\Gamma(3/\alpha)/\Gamma(1/\alpha)]^{-1/2}$, $\alpha > 0$ is the exponential rate of decay and σ^2 is distribution variance. Note that for $\alpha = 2$ the above is the Gaussian p.d.f. For $\alpha = 1$, it is the Laplacian p.d.f. Any i.i.d. source with distribution given by (7) is referred to as a generalized Gaussian source.

Numerical results for binary Markov channels with $\delta = 10$ and $M = 1$ and generalized Gaussian sources with shape parameter $\alpha = 0.5$, 1 and 2 are presented in Tables 1, 2 and 3, respectively. Signal-to-noise ratio (SNR) performances are given in dB for rates R=3 and 4 bits/sample and channel BER ϵ=0.005, 0.01, 0.05 and 0.1. Also provided in Tables 1-3 are the optimal performances theoretically attainable (OPTA) obtained by evaluating $D(RC)$, where $D(\cdot)$ is the distortion-rate function of the source for the squared-error distortion measure.

The COVQ results were obtained from 500,000 training vectors. A vector quantization codebook (optimized for the noiseless channel) with codewords assigned by a simulated annealing algorithm (described in [15]) is chosen as the initial codebook for the COVQ with ϵ=0.005. The final codebook for ϵ=0.005 is chosen as the initial codebook for ϵ=0.01, and so on.

The SQ-MAP results were obtained via simulations. The simulations were run 100 times, with N=1000 source samples used in each run. The average distortion, averaged over the 100 runs, is given in dB. The SQ's used in the simulations were symmetric Lloyd-Max scalar quantizers. As mentioned earlier, how the quantization levels are mapped to binary codewords is an important consideration. We have examined two codeword assignments: the natural binary code (NBC) and the folded binary code (FBC). An example of these two codes is illustrated in Figure 3. Note that the least significant bit (LSB) is the leftmost bit. Also, the FBC sign bit is the LSB. From our observations, FBC consistently outperforms NBC. FBC was used in the SQ-MAP results in Tables 1, 2 and 3.

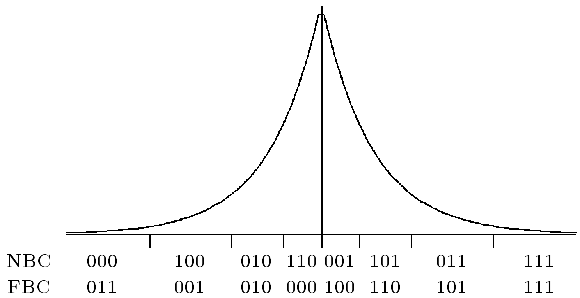

Figure 3. NBC and FBC Codeword Assignments for an 8-Level Lloyd-Max Scalar Quantizer; Generalized Gaussian Source with Shape Parameter $\alpha = 1$.

Note that, when M=1, $\tilde{Q}(\mathbf{e}_i|\mathbf{e}_{i-1})$ depends only on $\mathbf{e}_i$ and $e_{n(i-1)}$ (most significant bit (MSB) of $\mathbf{e}_{i-1}$). Thus, for fixed $\mathbf{y}^N$, the path metric from state $\mathbf{x}_{i-1}$ to state $\mathbf{x}_i$ depends only on $\mathbf{x}_i$ and $x_{n(i-1)}$ (MSB of $\mathbf{x}_{i-1}$). Therefore, the MSB of the binary codeword plays an important role in the Viterbi search. Now note that, because of symmetry, the MSB of NBC is 0 or 1 with equal probability. Hence, the MSB of NBC has zero redundancy. FBC, on the other hand, has the property that the MSB is much more likely to be 0 than 1. Hence, the MSB of FBC has high redundancy. Therefore, it is easier to determine whether $e_{n(i-1)}$=0 or 1 with FBC than with NBC. We believe that this is the reason for the superiority of FBC over NBC in the SQ-MAP scheme. We next compare COVQ and SQ-MAP.

The COVQ system is a (locally) optimal system that efficiently exploits the intra-block memory. Both encoder and decoder of this system are optimal in the sense of minimizing the mean squared error. However, this system does not make any use of the inter-block memory. On the other hand, the SQ-MAP system, which exploits both memories, consists of a sub-optimal encoder and a MAP decoder that minimizes the error probability but *not* the mean squared error. For fixed M, the effect of the intra-block memory of the channel becomes more dominant as kR increases. Therefore, for large blocks of kR bits, the COVQ system outperforms the SQ-MAP system (e.g. for $k=1$, $R=4$ in Tables 1-3).

V. Comparisons with Interleaving

The traditional technique for handling a channel with memory is to use interleaving. In the following, we consider two interleaving schemes and compare their performances against COVQ and SQ-MAP. The first scheme, COVQ-IL, consists of a COVQ optimized for a BSC and an interleaver. It is assumed that the interleaving length is sufficiently large so that the combination of interleaver, Markov channel and de-interleaver is equivalent to a BSC. The SNR performances of this scheme are given in Tables 1-3. COVQ-IL is compared against COVQ (optimized for the Markov channel). Observe that in all cases — except for α=0.5, R=3, k=1 and ϵ =0.005 — COVQ outperforms COVQ-IL. The gain of COVQ over COVQ-IL is due to the fact that COVQ exploits the noise memory whereas COVQ-IL does not.

The second interleaving scheme, SQ-IL, consists of an SQ designed by the Lloyd-Max formulation and an interleaver. The SQ binary codewords are assigned by FBC. The argument here is that FBC is a good codeword assignment for BSC [16] and the purpose of the interleaver/de-interleaver is to convert the Markov channel into a BSC. The SNR results of SQ-IL are provided in Tables 1-3. This scheme is compared against SQ-MAP. Note that SQ-MAP beats SQ-IL in all cases. The gain of SQ-MAP over SQ-IL is due to the residual redundancy, ρ_D, of the SQ and the noise memory. In Table 4, we provide the numerical values of ρ_D for α=0.5, 1 and 2 and R=3 and 4. Note that ρ_D tends to be larger for small α (broad-tailed distributions). Also, observe that the improvement of SQ-MAP over SQ-IL is larger for larger ρ_D.

Finally, we note that the two interleaving schemes have large encoding and decoding delays (due to the interleaver and de-interleaver). The COVQ scheme only have a block delay of $k-1$ samples. The SQ-MAP scheme has the MAP detector delay.

VI. Conclusions

We considered joint source-channel coding for real-valued i.i.d. sources and binary Markov channel. Two schemes were considered, COVQ and SQ-MAP. COVQ outperforms SQ-MAP when kR is large. These schemes were compared against two interleaving schemes. In most cases, the proposed schemes beat the interleaving schemes. In some instances, the performance gain is as much as 5 dB. These results, however, are still far from the theoretical limit (OPTA).

REFERENCES

[1] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, pp. 379–423 and 623–656, 1948.

[2] C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," *IRE Nat. Conv. Rec.*, pp. 142–163, Mar. 1959.

[3] E. Ayanoglu and R. M. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," *IEEE Transactions on Information Theory*, Vol. 33, pp. 855–865, Nov. 1987.

[4] N. Phamdo, N. Farvardin, and T. Moriya, "A Unified Approach to Tree-Structured and Multi-Stage Vector Quantization for Noisy Channels," *IEEE Transactions on Information Theory*, Vol. 39, pp. 835–850, May 1993.

[5] N. Phamdo and N. Farvardin, "Quantization Over Discrete Noisy Channels Using Rate-One Convolutional Codes," submitted to *IEEE Transactions on Information Theory*, Aug. 1993.

[6] F. Alajaji and T. Fuja, "A Communication Channel Modeled on Contagion," submitted to *IEEE Transactions on Information Theory*, Aug. 1993.

[7] H. Kumazawa, M. Kasahara, and T. Namekawa, "A Construction of Vector Quantizers for Noisy Channels," *Electronics and Engineering in Japan*, Vol. 67-B, pp. 39–47, Jan. 1984.

[8] N. Farvardin and V. Vaishampayan, "On the Performance and Complexity of Channel-Optimized Vector Quantizers," *IEEE Transactions on Information Theory*, Vol. 37, pp. 155–160, Jan. 1991.

[9] K. Sayood and J. C. Borkenhagen, "Use of Residual Redundancy in the Design of Joint Source/Channel Coders," *IEEE Transactions on Communications*, Vol. 39, pp. 838–846, June 1991.

[10] N. Phamdo and N. Farvardin, "Optimal Detection of Discrete Markov Sources Over Discrete Memoryless Channels — Applications to Combined Source-Channel Coding," to appear in *IEEE Transactions on Information Theory*, Jan. 1994.

[11] Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, Vol. 28, pp. 84–94, Jan. 1980.

[12] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory,* Vol. 28, pp. 129–137, Mar. 1982.

[13] J. Max, "Quantizing for Minimum Distortion," *IRE Transactions on Information Theory,* Vol. IT-6, pp. 7-12, Mar. 1960.

[14] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, "Detection of Binary Sources Over Discrete Channels with Additive Markov Noise," in preparation.

[15] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Transactions on Information Theory*, Vol. 36, pp. 799–809, July 1990.

[16] N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source-Channel Coding," *IEEE Transactions on Information Theory*, Vol. 33, pp. 827–838, Nov. 1987.

R	k	System	ϵ=0.005	ϵ=0.01	ϵ=0.05	ϵ=0.1
3	1	COSQ	7.80	7.75	6.52	5.23
		COSQ-IL	8.35	7.23	4.05	2.32
		SQ-MAP	9.27	8.37	5.67	4.03
		SQ-IL	7.94	6.35	0.78	-2.24
	2	COVQ	11.58	10.97	9.17	7.64
		COVQ-IL	10.99	9.65	5.73	3.75
	3	COVQ	13.59	12.72	10.35	9.11
		COVQ-IL	11.63	10.50	6.78	4.43
	∞	OPTA	21.59	21.46	20.63	19.76
4	1	COSQ	11.26	10.39	8.73	7.22
		COSQ-IL	10.79	9.09	5.39	3.38
		SQ-MAP	11.15	9.46	4.72	2.73
		SQ-IL	8.98	6.42	-0.46	-3.81
	2	COVQ	15.21	14.43	11.58	9.88
		COVQ-IL	13.75	12.11	7.48	4.94
	∞	OPTA	27.59	27.42	26.31	25.18

Table 1. SNR (in dB) Performances of Several Systems Operating Over a Markov Channel with $\delta = 10$ and $M = 1$; Generalized Gaussian Source with Shape Parameter $\alpha = 0.5$; R = Rate (Bits/Sample); k = Vector Dimension; ϵ = Channel Bit Error Rate; In the Interleaved Systems, COSQ and COVQ are Designed for Memoryless Channels; OPTA = Optimal Performance Theoretically Attainable.

R	k	System	ϵ=0.005	ϵ=0.01	ϵ=0.05	ϵ=0.1
3	1	COSQ	10.50	9.45	8.27	7.17
		COSQ-IL	10.49	9.17	5.21	3.62
		SQ-MAP	10.78	9.75	6.01	3.98
		SQ-IL	10.36	8.87	3.61	0.82
	2	COVQ	11.88	10.97	10.00	8.64
		COVQ-IL	11.67	10.28	6.60	4.47
	3	COVQ	13.01	12.43	10.68	9.48
		COVQ-IL	11.52	10.67	7.08	4.84
	∞	OPTA	18.54	18.42	17.59	16.74
4	1	COSQ	13.57	13.19	10.54	8.61
		COSQ-IL	12.76	11.03	6.82	4.79
		SQ-MAP	12.90	10.86	5.73	3.43
		SQ-IL	12.13	9.69	3.13	0.07
	2	COVQ	15.38	15.09	12.27	10.68
		COVQ-IL	14.41	12.92	8.33	5.71
	∞	OPTA	24.51	24.35	23.24	22.10

Table 2. SNR (in dB) Performances of Several Systems Operating Over a Markov Channel with $\delta = 10$ and $M = 1$; Generalized Gaussian Source with Shape Parameter $\alpha = 1$; R = Rate (Bits/Sample); k = Vector Dimension; ϵ = Channel Bit Error Rate; In the Interleaved Systems, COSQ and COVQ are Designed for Memoryless Channels; OPTA = Optimal Performance Theoretically Attainable.

R	k	System	ϵ=0.005	ϵ=0.01	ϵ=0.05	ϵ=0.1
3	1	COSQ	12.39	11.17	9.29	7.47
		COSQ-IL	12.04	10.50	6.47	4.67
		SQ-MAP	12.20	10.91	5.77	3.35
		SQ-IL	11.99	10.36	4.93	2.18
	2	COVQ	12.81	11.89	10.59	9.42
		COVQ-IL	12.46	11.15	7.36	5.15
	3	COVQ	13.55	12.90	11.35	10.05
		COVQ-IL	12.01	11.40	7.67	5.37
	∞	OPTA	17.92	17.80	16.96	16.11
4	1	COSQ	15.67	14.93	11.24	9.13
		COSQ-IL	14.15	12.30	7.81	5.60
		SQ-MAP	13.90	11.89	5.58	2.93
		SQ-IL	13.84	11.36	4.84	1.89
	2	COVQ	16.70	16.11	13.28	11.52
		COVQ-IL	15.28	13.70	9.06	6.40
	∞	OPTA	23.89	23.73	22.61	21.48

Table 3. SNR (in dB) Performances of Several Systems Operating Over a Markov Channel with $\delta = 10$ and $M = 1$; Generalized Gaussian Source with Shape Parameter $\alpha = 2$; R = Rate (Bits/Sample); k = Vector Dimension; ϵ = Channel Bit Error Rate; In the Interleaved Systems, COSQ and COVQ are Designed for Memoryless Channels; OPTA = Optimal Performance Theoretically Attainable.

α	R	$\rho_D = R - H(X)$
0.5	3	0.88
	4	1.05
1	3	0.42
	4	0.50
2	3	0.18
	4	0.23

Table 4. Redundancy (in Bits/Sample) of Symmetric Lloyd-Max Scalar Quantizer Output; Generalized Gaussian Source with Shape Parameter α; R = Rate of Scalar Quantizer in Bits/Sample.