Transactions Papers

Reconstruction of Predictively Encoded Signals Over Noisy Channels Using a Sequence MMSE Decoder

Farshad Lahouti, *Member, IEEE,* and Amir K. Khandani, *Member, IEEE*

***Abstract*—In this paper, we consider the problem of decoding a predictively encoded signal over a noisy channel when there is a residual redundancy (captured by a γ-order Markov model) in the sequence of transmitted data. Our objective is to minimize the mean-squared error (MSE) in the reconstruction of the original signal (input to the predictive source coder). The problem is formulated and solved through minimum mean-squared error (MMSE) decoding of a sequence of samples over a memoryless noisy channel. The related previous works include several maximum *a posteriori* (MAP) and MMSE-based decoders. The MAP-based approaches are suboptimal when the performance criterion is the MSE. On the other hand, the previously known MMSE-based approaches are suboptimal, since they are designed to efficiently reconstruct the data samples received (the prediction residues) rather than the original signal. The proposed scheme is set up by modeling the source-coder-produced symbols and their redundancy with a trellis structure. Methods are presented to optimize the solutions in terms of complexity. Numerical results and comparisons are provided, which demonstrate the effectiveness of the proposed techniques.**

***Index Terms*—Differential pulse code modulation (DPCM), forward–backward recursion, joint source-channel coding, Markov sources, maximum *a posteriori* (MAP) detection, minimum mean-squared error (MMSE) estimation, predictive quantization, residual redundancies, source decoding.**

I. Introduction

MOTIVATED by the fundamental work of Shannon [1], researchers have performed enormous endeavors on separate treatment of source and channel coders. However, in practice, due to strict design constraints such as restricted delay and limitations on the complexity of the systems involved, the joint design of source and channel coders has found increasing interest. Several paths have been taken toward the joint design of source and channel coders in the literature. These methods include optimized rate allocation, unequal error protection, optimized index assignment, channel-optimized quantization, and more recently, exploiting the source residual redundancies. Examples of papers on these topics can be found in [2]–[6] and [13]–[27]. For a comprehensive review of these techniques, the interested reader is referred to [7]–[12].

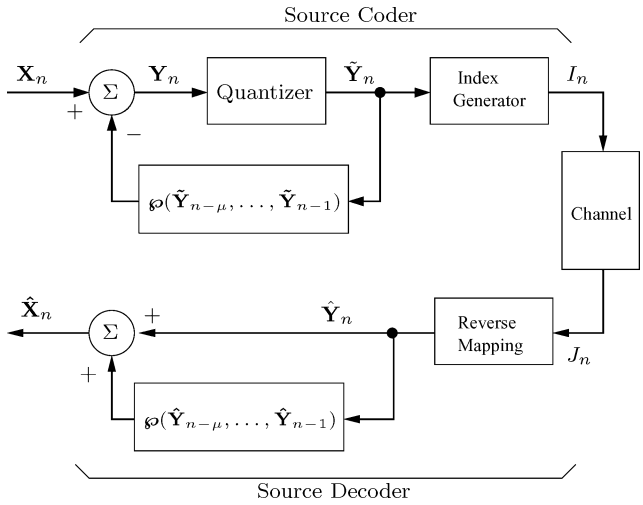

Fig. 1. Conventional differential pulse code modulation (DPCM) encoder and decoder with a (linear or nonlinear) MA prediction.

The work presented in this paper falls into the category of joint source-channel coders which use the *residual redundancy* [13] in the output of the source coder for improved reconstruction over noisy channels. This redundancy is due to the suboptimal source coding which is caused by, e.g., a constraint on complexity or delay. In general, this redundancy can be used for enhanced channel decoding, e.g., [14]–[18] or for effective source decoding, e.g., [19]–[24]. This is formulated in the form of a maximum *a posteriori* (MAP) detection or a minimum mean-squared error (MMSE) estimation problem. The residual redundancy is used both at the source and channel decoders in [25], which demonstrates an improved performance. In the same direction, iterative source and channel decoding schemes are presented in [26] and [27].

This paper considers the problem of reconstruction of a predictively quantized signal over a noisy channel when there is a residual redundancy in the source-coder output stream. In fact, it is shown in [13] that there is always a residual redundancy in the output of a predictive quantizer, due to a mismatch between the encoder prediction model and that of the source.

Paper approved by K. Rose, the Editor for Source-Channel Coding of the IEEE Communications Society. Manuscript received September 18, 2002; revised October 16, 2003 and December 19, 2003. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada. This paper was presented in part at the 39th Annual Allerton Conference on Communication, Control, and Computing, October 2001.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: farshad@cst.uwaterloo.ca; khandani@cst.uwaterloo.ca).

Digital Object Identifier 10.1109/TCOMM.2004.833028

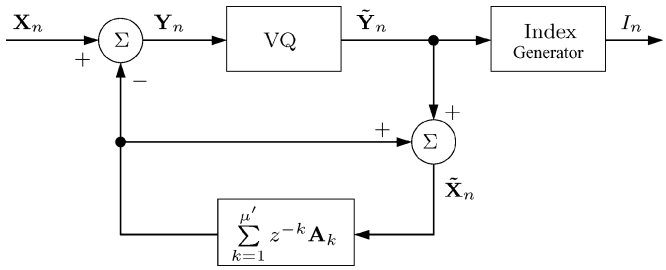

Fig. 2. DPCM encoder with linear AR prediction.

In predictive coding schemes, the signal that is quantized, $\mathbf{Y}_n$, is the *prediction residue*, or the difference between the original signal $\mathbf{X}_n$ and its estimate, produced using a prediction function. In moving-average (MA) systems (see Fig. 1), the prediction function operates based on μ previous quantized prediction residues $(\tilde{\mathbf{Y}}_{n-\mu}, \ldots, \tilde{\mathbf{Y}}_{n-1})$, whereas in autoregressive (AR) systems (see Fig. 2), μ' previous quantized signals $(\tilde{\mathbf{X}}_{n-\mu'}, \ldots, \tilde{\mathbf{X}}_{n-1})$ are used for prediction. Alternatively, autoregressive moving-average (ARMA) systems use both sets of data.[1] For a comprehensive review of predictive quantization, refer to [28] and [29].

The output of a predictive coder is the index I_n corresponding to the quantized prediction residue $\tilde{\mathbf{Y}}_n$. In a basic predictive decoder, the block labeled "reverse mapping" in Fig. 1 is simply the inverse of the index-generation function at the encoder, which ignores any residual redundancy in the source-coder output. Recently, researchers have replaced this block with more sophisticated systems which exploit the residual redundancy for improved reconstruction. Sayood and Borkenhagen in [13] proposed a MAP-based decoder. In [21], for a differential pulse code modulation (DPCM)-encoded speech, an MMSE-based scheme is employed that aims at minimizing the error in reconstructing the prediction residue $\mathbf{Y}_n$ at the receiver. Both schemes of [13] and [21] use a first-order Markov model to capture the residual redundancy. For reconstruction of a DPCM-encoded image, several schemes have been suggested which exploit the residual redundancy, both in the horizontal and the vertical directions [30]–[32]. A scheme called maximal signal-to-noise ratio (SNR) decoding is suggested in [30], which searches for the residue codeword that minimizes a simplified expression of the reconstructed signal SNR. This simplification reduces the objective function to one that represents the error in the reconstruction of the prediction residues. Another scheme for the reconstruction of DPCM-encoded images is the MMSE-based decoder proposed in [31], which uses a Markov mesh for the reconstruction of the prediction residues. In general, the MAP-based approaches are suboptimal when the performance criterion is the mean-squared error (MSE) and the previously known MMSE-based approaches are suboptimal, since they aim at minimizing the error in reconstruction of the prediction residues, rather than the original signal (input to the source coder).

In this paper, our objective is to design a source decoder (not a reverse-mapping unit) which minimizes the MSE in the reconstruction of the original signal, when the residual redundancy is captured by a γ-order Markov model ($\gamma \geq 1$) and a delay of δ ($\delta \geq 0$) is allowed in the decoding process. The problem is formulated and solved through MMSE decoding of a sequence of samples over a memoryless noisy channel, which was previously recognized to be an open problem by Phamdo and Farvardin [19]. The solution is set up by modeling the stream of encoder-produced symbols and their redundancy with a trellis structure. The proposed solution is optimized to minimize the computational complexity.

The organization of the paper is as follows. The notations, system, and channel model used are presented in Section II. In Section III, the sequence MMSE (SMMSE) decoder is presented. In Section IV, the application of the proposed SMMSE decoding scheme for the reconstruction of an AR DPCM-encoded source is discussed. The systems considered for comparison, numerical results, and analysis are presented in Section V.

II. Preliminaries

A. Notations

The notations used in this paper are as follows. The capital letters, e.g., I, represent random variables, while the small letters, e.g., i, represent a realization. We replace the probability $P(I = i)$ by $P(I)$ in most instances, when it does not lead to confusion. The vectors are shown boldfaced, e.g., $\mathbf{X}$. The lower index indicates the time instant, e.g., $\mathbf{X}_n$ is the vector $\mathbf{X}$ at time instant n. The upper index in parenthesis indicates components of a vector or bit positions of an integer, e.g., $\mathbf{X}_n = [X_n^{(1)}, \ldots, X_n^{(N)}]$, where N is the dimension of the vector $\mathbf{X}_n$. A sequence of variables over time, e.g., $(I_{n_1}, \ldots, I_{n_2})$, $n_1 \leq n_2$ is denoted by $\underline{I}_{n_2}^{n_1}$. For simplicity, we represent $\underline{I}_n^1$ by $\underline{I}_n$. The N-dimensional Cartesian product of a set $\mathcal{J}$ is represented by $\mathcal{J}^N$, which consists of N-dimensional vectors whose components are taken from $\mathcal{J}$.

B. System Overview

The block diagram of the system is shown in Fig. 1. The source coder is a mapping from an N-dimensional Euclidean space $\mathcal{R}^N$ into a finite index set $\mathcal{J}$ of M elements. It is composed of two components: a predictive quantizer and an index generator. The predictive quantizer maps the input sample $\mathbf{X} \in \mathcal{R}^N$ to one of the reconstruction points or *codewords* in the codebook $\mathcal{C} \subset \mathcal{R}^N$. The example of a predictive quantizer shown in Fig. 1 uses an MA prediction system. The index generator then maps the codeword selected by the quantizer to an *index (symbol)* I in the index set $\mathcal{J}$. The bitrate of the quantizer is given by $r = \lceil \log_2 M \rceil$ b/symbol (or $\lceil \log_2 M \rceil / N$ b/dim).

We assume that the quantized sample $\tilde{\mathbf{X}}_n$ corresponding to the predictive quantizer input $\mathbf{X}_n$ can be described as a function $\mathbf{f}$ of the last $\mu + 1$ encoded symbols, i.e.,

$$\tilde{\mathbf{X}}_n = \mathbf{f}(I_{n-\mu}, \ldots, I_{n-1}, I_n), \quad I_{n-k} \in \mathcal{J}, \quad 0 \leq k \leq \mu \tag{1}$$

where μ denotes the memory length of the predictor. A concrete example is the MA predictive quantizer of Fig. 1, for which we have

$$\tilde{\mathbf{X}}_n = \tilde{\mathbf{Y}}_n + \wp(\tilde{\mathbf{Y}}_{n-\mu}, \ldots, \tilde{\mathbf{Y}}_{n-1}) \tag{2}$$

[1] In this paper, the terms AR, MA, and ARMA predictive systems indicate using either a linear or nonlinear predictive function, unless specifically specified.

and noting that the index generator is a simple one-to-one mapping function, (1) holds. Note that in the absence of channel errors $\hat{\mathbf{X}}_n = \mathbf{X}_n$ and $\hat{\mathbf{Y}}_n = \mathbf{Y}_n$. In Section IV, we demonstrate that a DPCM scheme with an AR predictor can also be cast into the model of (1) with approximation.

At the receiver, for each transmitted r-bit index $I = i$, a vector J with r components is received, which provides information about I. At time instant n, the reconstructor (source decoder) produces the output signal $\hat{\mathbf{X}}_n$. In this reconstruction, the source decoder uses J_n, and it may also use the previously received samples and/or some of the future samples.

C. Channel Model

The channels considered in this paper are described by a probability density function (pdf) $P(J_n|I_n)$. We assume that the channel is memoryless without intersymbol interference (ISI) in the sense that, for a sequence of transmitted symbols $\underline{I}_n = (I_1, I_2, \ldots, I_n)$ and the corresponding received signals $\underline{J}_n$, the following equality is valid:

$$P(J_n = j_n|\underline{I}_n = \underline{i}_n, \underline{J}_{n-1} = \underline{j}_{n-1}) = P(J_n = j_n|I_n = i_n). \tag{3}$$

This results in the following:

$$P(J_n = j_n|\underline{I}_n = \underline{i}_n) = P(J_n = j_n|I_n = i_n) \tag{4}$$

$$P(\underline{J}_n = \underline{j}_n|\underline{I}_n = \underline{i}_n) = \prod_{k=1}^{n} P(J_k = j_k|I_k = i_k). \tag{5}$$

An example is a binary phase-shift keying (BPSK) modulation over a channel with additive white Gaussian noise (AWGN) which produces soft outputs as

$$j_n^{(m)} = s\left(i_n^{(m)}\right) + \eta_n^{(m)}, \quad m = 1, \ldots, r \tag{6}$$

where $i_n^{(m)}, m = 1, \ldots, r$ are the bit components of i_n or the source-coder output and $j_n^{(m)}$ are the corresponding channel soft outputs, $\boldsymbol{\eta}_n = [\eta_n^{(1)}, \ldots, \eta_n^{(r)}]$ is a vector of independent and identically distributed (i.i.d.) Gaussian noise samples, and $s(.) \in \{\sqrt{E_b}, -\sqrt{E_b}\}$ is a mapping of bits to channel signals. The relationship between the transmitted and the received symbols is then given by the following conditional pdf:

$$P(J_n = j_n|I_n = i_n) = \prod_{m=1}^{r} P(j_n^{(m)}|i_n^{(m)}). \tag{7}$$

In this paper, we refer to such a channel as the soft output channel (SOC) model. The binary symmetric channel (BSC) model is also based on (6), when a hard decision is made on the received soft outputs. If the resulting bit-error probability (BEP) is denoted by ϵ, then the relationship between the transmitted and the received symbols is given by

$$P(J_n = j_n|I_n = i_n) = (\epsilon)^{h(i_n, j_n)}(1 - \epsilon)^{r - h(i_n, j_n)} \tag{8}$$

where j_n is the received binary codeword in $\mathcal{J}$ and $h(i_n, j_n)$ is the Hamming distance between indexes i_n and j_n. In the following, for the development of the proposed source decoders, we assume that the probability distribution of $P(J_n|I_n)$ is given, and the memoryless channel assumption of (3) is valid.

III. MMSE Source Decoding

Consider the case where due to the suboptimality of the predictive source coder, there is a residual redundancy in its output stream. This redundancy is in the form of a memory in the sequence of the transmitted symbols or also in the form of a nonuniform symbol probability distribution. Our objective is to design a source decoder that exploits this residual redundancy to effectively reconstruct the original source samples at the receiver. We begin by reviewing the asymptotically optimum MMSE (AOMMSE) decoder originally presented in [11] for memoryless source coders. Specifically, we observe that the same formulation is applicable to the case of predictive source coders, as well. Next, in Section III-B, we present a SMMSE decoder.

A. AOMMSE Decoder

In the context of MMSE decoding, a source decoder is designed to minimize

$$E[(\mathbf{X}_n - \hat{\mathbf{X}}_n)'(\mathbf{X}_n - \hat{\mathbf{X}}_n)|\underline{J}_{n+\delta}] \tag{9}$$

in which $\mathbf{X}_n$ is the source sample, and $\hat{\mathbf{X}}_n$ is its estimate, given the received sequence $\underline{J}_{n+\delta} = [J_1, J_2, \ldots, J_{n+\delta}]$. In (9), $\delta \geq 0$ is the delay allowed in the decoding process. Based on the fundamental theorem of estimation, the optimum solution minimizing (9) is given by

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n|\underline{J}_{n+\delta}] \tag{10}$$

that is expanded to

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_{n+\delta} \in \mathcal{J}^{n+\delta}} E[\mathbf{X}_n|\underline{I}_{n+\delta}] P(\underline{I}_{n+\delta}|\underline{J}_{n+\delta}). \tag{11}$$

In (11), $E[\mathbf{X}_n|\underline{I}_{n+\delta}]$ forms the *decoder codebook*. Therefore, (11) presents an optimal decoder that at time n requires a sum over $M^{n+\delta}$ elements of the decoder codebook. In this case, both computational complexity and the memory requirement grow exponentially with time, leading to an impractical scheme. Assuming that the source $\mathbf{X}$ has a memory that asymptotically decays with time, for sufficiently large values of $\tau, \tau \in \mathcal{Z}$, the decoder codebook can be approximated by

$$E[\mathbf{X}_n|\underline{I}_{n+\delta}] \approx E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-\tau}] \tag{12}$$

and therefore, the MMSE decoder given by

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_{n+\delta}^{n-\tau}} E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-\tau}] P(\underline{I}_{n+\delta}^{n-\tau}|\underline{J}_{n+\delta}) \tag{13}$$

is asymptotically optimal, and yet feasible for limited values of τ. This decoder is, in fact, the AOMMSE decoder derived in [11] for a memoryless source coder. It is clear that the same formulation is applicable to the case of a predictive coder, and is asymptotically optimum in the MMSE sense for sufficiently large values of τ. In this case, the decoder codewords $E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-\tau}]$ provide a finer reconstruction of the source samples as compared with the quantized signal at the encoder (given by (1) as a function of only $\underline{I}_n^{n-\mu}$).

For MMSE reconstruction of sources encoded with a memoryless source coder, several interesting works are reported in [19], [20], and [36]. All these papers consider a first-order

Markov model for the residual redundancy. For $\gamma = 1$, $\tau = \delta = 0$, the decoder in (13) is simplified to the instantaneous MMSE decoder of [19]. If $\gamma = \tau = 1$, $\delta \geq 0$, then the AOMMSE decoder is equivalent to the Markov type-1 decoder of [36] used for a memoryless channel. The sequence-based MMSE decoder of [20] uses the decoder codebook $E[\mathbf{X}_n|I_n]$ or $E[\mathbf{X}_n|I_n, I_{n-1}]$ and accommodates delay in the decoding process. In fact, the "memory-enhanced" decoder of [20] is given by

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_n^{n-1}} E[\mathbf{X}_n|\underline{I}_n^{n-1}]P(\underline{I}_n^{n-1}|\underline{J}_{n+\delta}).$$

B. SMMSE Decoder

Now, we turn our attention to derive a simplified MMSE decoder for predictive source coders. Specifically, we are interested in a source decoder which uses a decoder codebook similar to its corresponding encoder (quantization) codebook. This is of particular interest, since it leads to a less complex decoder with a significantly smaller memory requirement, especially in symmetric communication systems where the encoder codebook is already available at the receiver. Consequently, we consider the following SMMSE decoder:

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_n^{n-\mu} \in \mathcal{J}^{\mu+1}} \mathbf{f}\left(\underline{I}_n^{n-\mu}\right) P(\underline{I}_n^{n-\mu}|\underline{J}_{n+\delta}) \tag{14}$$

which provides the MMSE estimate as a weighted average of the reconstruction values $\mathbf{f}(\underline{I}_n^{n-\mu})$ [see (1)]. Each weight or the probability $P(\underline{I}_n^{n-\mu} = \underline{i}_n^{n-\mu}|\underline{J}_{n+\delta})$ is the *a posteriori* probability (APP) of a sequence of symbols calculated in every time instant. We do not assume any restriction on the definition of the function $\mathbf{f}(\cdot)$ at this time. As described in Section II-B, it particularly reflects the fact that the source coder has memory, and as a result, the decoded signal is a function of a set of encoded symbols. Depending on the choice of the source coder, this function is determined accordingly. In the following section, methodologies are presented to calculate the required APPs. Subsequently, in Section IV, a decoder for reconstruction of an AR DPCM-coded signal over a noisy channel, based on the proposed SMMSE decoder, is presented.

C. Calculating the Weights

To calculate the APPs required in the proposed SMMSE decoder of (14), we assume that the encoder output symbols form a γ-order Markov model due to the residual redundancies. These symbols are then modeled by a trellis structure. In this structure, the states at time n are defined by the ordered set

$$S_n = (I_{n-\gamma+1}, I_{n-\gamma+2}, \ldots, I_{n-1}, I_n), \quad I_{n-k} \in \mathcal{J}, \quad 0 \leq k < \gamma. \tag{15}$$

Hence, there are M^γ states in each time step (stage), $S_n \in \mathcal{J}^\gamma$. Each branch leaving the state at time step n corresponds to one particular symbol $I_{n+1} = i_{n+1}$. Therefore, there are M branches leaving each state. Each branch is identified by the pair $(S_n = s_n, S_{n+1} = s_{n+1})$ of the two states that the branch connects together. Having defined the trellis structure as such, there will be one *a priori* probability $P(I_{n+1} = i_{n+1}|S_n = s_n)$ corresponding to each branch which characterizes the γ-order Markov property of the source-coder symbols. The states now form a first-order Markov sequence. Using this property and the memoryless assumption of the channel [see (3)–(5)], in line with the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm [33], the probability of a particular state S_n, given the observed sequence $\underline{J}_{n+\delta}$, is calculated recursively by the following forward–backward equation:

$$P(S_n|\underline{J}_{n+\delta}) = C \cdot P(S_n|\underline{J}_n) \cdot P(\underline{J}_{n+\delta}^{n+1}|S_n) \tag{16}$$

where C is a factor which normalizes the sum of the probabilities to one. The term $P(S_n|\underline{J}_n)$ is the forward term, and is given by

$$P(S_n|\underline{J}_n) = C \cdot P(J_n|I_n) \cdot \sum_{S_{n-1} \to S_n} P(I_n|S_{n-1})P(S_{n-1}|\underline{J}_{n-1}) \tag{17}$$

where the summation is over a subset of M states in time step $n-1$, which are connected to the state S_n. The term $P(\underline{J}_{n+\delta}^{n+1}|S_n)$ in (16) is the backward term, and is calculated recursively by

$$P(\underline{J}_{n+\delta}^{n+1}|S_n) = \sum_{I_{n+1} \in \mathcal{J}} P(J_{n+1}|I_{n+1}) \cdot P(I_{n+1}|S_n) \cdot P(\underline{J}_{n+\delta}^{n+2}|S_{n+1}). \tag{18}$$

In (18), the term $P(\underline{J}_{n+\delta}^{n+2}|S_{n+1})$ is the left-hand side (LHS) of the same equation for time instant $n+1$. This creates a recursion that is computed backward in time, and starts from

$$P(\underline{J}_{n+\delta}^{n+\delta}|S_{n+\delta-1}) = P(J_{n+\delta}|S_{n+\delta-1}) = \sum_{I_{n+\delta} \in \mathcal{J}} P(J_{n+\delta}|I_{n+\delta}) \cdot P(I_{n+\delta}|S_{n+\delta-1}). \tag{19}$$

We note that in each time step, the forward recursion of (17) proceeds one step forward through the trellis, while the backward term is recomputed over the entire backward window, as indicated in (18) and (19). The details of the derivation of these equations are provided in the Appendix.[2]

The presented trellis structure and the forward and backward equations are used in the following sections for calculation of the required probabilities (weights) in (14). Depending on the relative value of encoder memory μ to the residual redundancy order γ, this is performed in two ways, as described below.

1) Calculating the Weights for $\mu < \gamma$: For the scenario with $\mu < \gamma$, we can calculate the probabilities required in (14), by performing $\gamma - \mu - 1$ summations over *any of the state probabilities* $P(S_{n+m}|\underline{J}_{n+\delta})$ *as long as* S_{n+m} *includes* $\underline{I}_n^{n-\mu}$, *or equivalently,* $0 \leq m \leq \gamma - \mu - 1$. However, it is straightforward to see that the number of computations required for the forward

[2]It is noteworthy that the forward–backward algorithm [33] has been used in different forms and applications, such as channel decoding and the decoding in hidden Markov models (HMM). In the context of joint source-channel coding, similar developments are related to the BCJR algorithm [33] and some other classic works in [19], computations over HMMs in [20], or the prediction and filtering within the context of Kalman filtering in [36]. The sum-product algorithm [37] presents a general mathematical framework that contains these separately known algorithms, and shows their relationships. Our focus here is to investigate the details of the calculation of the specific APPs required, so as to find more efficient solutions.

and backward recursions per time step (denoted by NC_{fwd} and NC_{bwd}, respectively) is given by

$$NC_{\text{fwd}} = (2M+3)\,M^{\gamma} \tag{20}$$
$$NC_{\text{bwd}} = 3(\delta - m)\,M^{\gamma+1} \tag{21}$$

where $\delta - m$ is the number of backward recursions required per time step. Therefore, we can select the value of m such that it minimizes the overall computational burden, which consists of the computations required for the forward and the backward terms. Noting that only NC_{bwd} depends on m, we solve the following for the optimum value of m:

$$\begin{aligned}&\text{Minimize } NC_{\text{bwd}} = 3(\delta - m)\cdot M^{\gamma+1}\\ &\text{subject to } 0 \le m \le \gamma - \mu - 1, \quad 0 \le m \le \delta.\end{aligned} \tag{22}$$

Case 1. $\delta < \gamma - \mu$: In the cases where the delay is smaller than the difference of the assumed residual redundancy order and the encoder memory, we are able to choose $m = \delta$ and eliminate the backward term. The probabilities in (14) are calculated using (17) and the following:

$$P(\underline{I}_n^{n-\mu}|\underline{J}_{n+\delta}) = \cdots \sum_{I_{n+k}} \cdots P(S_{n+\delta}|\underline{J}_{n+\delta}),$$
$$k = \delta - \gamma + 1, \ldots, \delta, \quad k \neq -\mu, \ldots, 1, 0 \tag{23}$$

where (23) indicates $\gamma - \mu - 1$ summations over the probabilities of states at time step $n+\delta$, $S_{n+\delta} = (I_{n-\gamma+\delta+1}, \ldots, I_{n+\delta})$.

Case 2. $\delta \ge \gamma - \mu$: Alternatively, when the delay is larger than $\gamma - \mu$, the NC_{bwd} is minimized when $m = \gamma - \mu - 1$, i.e., $\delta + \mu - \gamma + 1$ backward recursions are required. The probabilities in (14) are now given by

$$P(\underline{I}_n^{n-\mu}|\underline{J}_{n+\delta}) = \sum_{I_{n+1}} \sum_{I_{n+2}} \cdots \sum_{I_{n+\gamma-\mu-1}} P(S_{n+\gamma-\mu-1}|\underline{J}_{n+\delta}) \tag{24}$$

and (17)–(19).

Note that in both cases described above, the computational complexity remains $O(M^{\gamma+1})$ complex, but is optimized (reduced) by multiples of $M^{\gamma+1}$ depending on the system parameters. The dynamic memory required is of the order of $O(M^{\gamma})$, which corresponds to the number of states. This is also the case for the static memory required to store the *a priori* probabilities. The weight computations for the instantaneous MMSE decoder of [19], corresponds to *Case 1* above, with $\mu = 0, \gamma = 1, \delta = 0$, which indicates $2M^2 + 3M$ computations. The weight computations required for the decoder of [20] corresponds to *Case 2* above, with $\mu = 0, \gamma = 1, \delta > 0$, which indicates $m = 0$, and $(2 + 3\delta)M^2 + 3M$ computations.

2) Calculating the Weights for $\mu \ge \gamma$: For the scenario with the residual redundancy order smaller than the encoder memory $\mu \ge \gamma$, the sequence $\underline{I}_n^{n-\mu} = (I_{n-\mu}, \ldots, I_{n-1}, I_n)$ whose APP is required, in fact corresponds to a sequence of states within the trellis structure of the source-coder-produced symbols, as described before. Consequently, the desired probabilities can be calculated using the probability of the corresponding sequence of states. We have

$$P(\underline{I}_n^{n-\mu}|\underline{J}_{n+\delta}) = P(\underline{S}_n^{n-\mu+\gamma-1}|\underline{J}_{n+\delta}). \tag{25}$$

This can be written in the following forward–backward form where we have used the assumption of redundancy order of γ, to replace $P(\underline{J}_{n+\delta}^{n+1}|\underline{S}_n^{n-\mu+\gamma-1})$ with $P(\underline{J}_{n+\delta}^{n+1}|S_n)$:

$$P(\underline{S}_n^{n-\mu+\gamma-1}|\underline{J}_{n+\delta}) = C \cdot P(\underline{S}_n^{n-\mu+\gamma-1}|\underline{J}_n) \cdot P(\underline{J}_{n+\delta}^{n+1}|S_n). \tag{26}$$

The value C is a factor which normalizes the sum of probabilities to one. The second term or the backward term is given by (18) and (19). The forward term is given by

$$\begin{aligned}P(\underline{S}_n^{n-\mu+\gamma-1}|\underline{J}_n) &= \left[\prod_{k=-\mu+\gamma}^{0} P(J_{n+k}|I_{n+k}) \cdot P(I_{n+k}|S_{n+k-1})\right]\\ &\quad \cdot P(S_{n-\mu+\gamma-1}|\underline{J}_{n-\mu+\gamma-1}).\end{aligned} \tag{27}$$

The computations to find the APPs $P(\underline{I}_n^{n-\mu}|\underline{J}_{n+\delta})$ for the SMMSE decoder of (14), as described above for the case of $\mu \ge \gamma$, is $O(M^{\mu+1})$ complex and requires a dynamic memory proportional to M^{γ} or the number of states. Alternative (less computationally complex) ways to calculate this probability are provided by using the extended trellis structure described in [11] and [12]. The general problem is stated as finding the APP of a sequence of $\mu + 1$ symbols of a γ-order Markov source transmitted over a noisy channel. In the extended trellis structure of [11], each state, referred to as a *super state*, is defined by a sequence of $L' + \gamma$ symbols, $0 \le L' \le \mu - \gamma$. It is demonstrated in [11] that by increasing L', the complexity of computations are decreased, at the cost of an increase in the memory requirements. Furthermore, it is shown that increasing L' beyond $\mu - \gamma$ results in solutions that are inefficient, both in terms of computational complexity and memory requirements [11]. Note that the specific case of $L' = \mu - \gamma + 1$ indicates a trellis, in which states correspond to the sequences $\underline{I}_n^{n-\mu} = \underline{i}_n^{n-\mu}$, that their APPs are to be computed.

IV. Reconstruction of Predictively Encoded Signals

In this section, we consider the MMSE reconstruction of a linear AR DPCM coded signal over a noisy channel. This focus is due to the popularity of these systems and the fact that the ideas employed in this case can be easily applied to the other cases, including MA (linear or nonlinear) predictive encoding systems.

Fig. 2 demonstrates the block diagram of a DPCM encoder with a linear AR prediction. In this system, the quantized sample $\tilde{\mathbf{X}}_n$ is given by

$$\tilde{\mathbf{X}}_n = \tilde{\mathbf{Y}}_n + \sum_{k=1}^{\mu'} \mathbf{A}_k \tilde{\mathbf{X}}_{n-k}. \tag{28}$$

By recursive replacement of $\tilde{\mathbf{X}}_{n-k}$ in (28), it is straightforward to see that $\tilde{\mathbf{X}}_n$ can be described as a function of the sequence of quantized prediction residues $(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_n)$. This follows the general form of (1), except that the length of the sequence to be decoded, $\underline{\tilde{\mathbf{Y}}}_n$, grows with time. A manageable solution is created by defining an *effective memory* length, i.e., assuming that the sample $\tilde{\mathbf{X}}_n$ depends *effectively* only on $\tilde{\mathbf{Y}}_n$ and μ previous prediction-residue values $(\tilde{\mathbf{Y}}_{n-\mu}, \ldots, \tilde{\mathbf{Y}}_{n-1}, \tilde{\mathbf{Y}}_n)$. Therefore,

we can finalize the reconstructed value of the residues beyond $n-\mu$, or equivalently, their corresponding output $\hat{\mathbf{X}}_{n-\mu-1}$. This idea is supported by the fact that in DPCM systems, error in one sample is effectively propagated to only a limited number of future samples. Using this concept, we now consider the case of a first-order AR DPCM system in more detail.

For a first-order AR predictive coder, we have

$$\mathbf{X}_n = \mathbf{Y}_n + \mathbf{A}\tilde{\mathbf{X}}_{n-1} \tag{29}$$

$$= \mathbf{Y}_n + \sum_{k=1}^{\mu} \mathbf{A}^k \tilde{\mathbf{Y}}_{n-k} + \mathbf{A}^{\mu+1}\tilde{\mathbf{X}}_{n-\mu-1} \tag{30}$$

$$= \mathbf{Z}_n + \mathbf{A}^{\mu+1}\tilde{\mathbf{X}}_{n-\mu-1} \tag{31}$$

where

$$\mathbf{Z}_n \triangleq \mathbf{Y}_n + \sum_{k=1}^{\mu} \mathbf{A}^k \tilde{\mathbf{Y}}_{n-k}. \tag{32}$$

Using (31), the MMSE estimate (10) is now given by

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n | \underline{J}_{n+\delta}]$$
$$= E[\mathbf{Z}_n | \underline{J}_{n+\delta}] + \mathbf{A}^{\mu+1} E[\tilde{\mathbf{X}}_{n-\mu-1} | \underline{J}_{n+\delta}]. \tag{33}$$

Subsequently, assuming an effective memory length of μ, we approximate the second term by $\mathbf{A}^{\mu+1}\hat{\mathbf{x}}_{n-\mu-1}$. Next, we reach a recursive formula for MMSE decoding of a first-order AR DPCM system

$$\hat{\mathbf{x}}_n = \hat{\mathbf{z}}_n + \mathbf{A}^{\mu+1}\hat{\mathbf{x}}_{n-\mu-1} \tag{34}$$

where

$$\hat{\mathbf{z}}_n = E[\mathbf{Z}_n | \underline{J}_{n+\delta}] \tag{35}$$

can be calculated using the AOMMSE decoder of (13) or the SMMSE decoder of (14). The latter is motivated by (32) and the fact that $\tilde{\mathbf{Y}}_n = E[\mathbf{Y}_n | I_n]$, which results in the following expression for the quantized value of $\mathbf{Z}_n$:

$$\tilde{\mathbf{Z}}_n = \sum_{k=0}^{\mu} \mathbf{A}^k E[\mathbf{Y}_{n-k} | I_{n-k}]$$
$$= \mathbf{f}(I_{n-\mu}, \ldots, I_{n-1}, I_n) \tag{36}$$

as required by the assumption of (1). Subsequently, computing (35) via (36) and (14), the solution based on the SMMSE decoder for reconstruction of a first-order DPCM encoded signal over a noisy channel is given by

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_n^{n-\mu}} \left[\sum_{k=0}^{\mu} \mathbf{A}^k E[\mathbf{Y}_{n-k} | I_{n-k}] \right] P(\underline{I}_n^{n-\mu} | \underline{J}_{n+\delta}) + \mathbf{A}^{\mu+1}\hat{\mathbf{x}}_{n-\mu-1} \tag{37}$$

in which the decoder codebook is determined by the encoder codebook. Note that the $E[\mathbf{Y}|I]$ in (36) and (37) is the encoder (quantization) codeword, assuming a Linde–Buzo–Gray (LBG) vector quantizer [35]. It is noteworthy that for $\mu = 0$, the solution collapses to that of the MMSE reconstruction of prediction residues

$$\hat{\mathbf{x}}_n = \sum_{I_n \in \mathcal{J}} E[\tilde{\mathbf{Y}}_n | I_n] P(I_n | \underline{J}_{n+\delta}) + \mathbf{A}\hat{\mathbf{x}}_{n-1}. \tag{38}$$

TABLE I
COEFFICIENTS OF SYNTHESIZED SOURCE A

coefficient (1-5)	1.1160	0.5365	-0.1830	-0.5205	-0.0535
coefficient (6-10)	-0.3159	0.3263	-0.0194	0.2841	-0.2006

TABLE II
REDUNDANCY OF SOURCE-CODER OUTPUT $R(M,\gamma)$ (BITS) AT DIFFERENT REDUNDANCY MODEL ORDERS γ ($M = 8, N = 1$)

Redundancy Order γ	0	1	2	3
$R_A(M,\gamma)$ (bits)	0.34	1.15	1.40	1.44
$R_B(M,\gamma)$ (bits)	0.20	0.36	0.43	0.46

In (37), the assumption of an effective memory length of μ reflects the fact that $\hat{\mathbf{x}}_n$ is composed of the decoded sample $\hat{\mathbf{x}}_{n-\mu-1}$ and the *soft* (undecoded) information on symbols $\underline{I}_n^{n-\mu}$, which are positioned within the effective memory of the predictive decoder. This soft information is encapsulated in APPs $P(\underline{I}_n^{n-\mu} | \underline{J}_{n+\delta})$. In Section V, we investigate the performance of this decoder. Note that the required probabilities of this decoder, $P(\underline{I}_n^{n-\mu} | \underline{J}_{n+\delta})$, are calculated using the different methods provided in Section III-C, depending on the relative values of γ, μ, and δ.

V. PERFORMANCE ANALYSIS

To analyze the performance of the proposed decoders, we use two synthesized sources similar to [13]. The first source, referred to as source A, is a tenth-order Gauss–Markov source with the coefficients given in Table I. The coefficients correspond to the linear predictive coding (LPC) coefficients of a 20-ms segment of speech. The second source, source B, is a first-order Gauss–Markov source with a correlation coefficient of 0.95. The source $\mathbf{X}_n = [X_{(n-1)N+1}, \ldots, X_{nN}]$ is the input to a first-order linear AR DPCM encoder. The quantizer is an M-point N-dimensional LBG vector quantizer (VQ) [35]. In our particular example, we consider $M = 8$, $N = 1$ (Lloyd–Max scalar quantizer), and a predictor designed for noisy channels [34]. The predictor coefficients for sources A and B are then selected as 0.833 and 0.724, respectively. The index assignment is natural binary or Gray coding, and we use a SOC model as described in Section II-C. A set of 1 000 000 samples is used for simulation (test), which is separate from the set of samples used to train the quantizer. For each decoder, the test data set is transmitted over ten different realizations of the noisy channel (random seeds) and decoded at the receiver. As a measure of decoding performance, the average reconstructed SNR is calculated. We also analyze the sensitivity of the decoding performance to different channel realizations.

Table II presents the value $R(M,\gamma)$ (in bits) defined as

$$R(M,\gamma) \triangleq \log_2 M - H(I_n | S_{n-1}) \tag{39}$$

where $S_n = (I_{n-\gamma+1}, \ldots, I_n)$ as an indication of the available redundancy at the output of the source coder, and hence, the gain to be achieved using different redundancy model orders γ. A similar expression up to a scaling for the case of a first-order Markov model is presented in [13] and referred to as the error-correction capability index. As given in this Table, for source A, the redundancy due to the nonuniform distribution

($\gamma = 0$) is 0.34 bits. The redundancy exploited by means of a first-, second- and third-order Markov model is 1.15, 1.40, and 1.44 b, respectively.

A. Systems for Comparison: ML, MMSE, SMAP, MSNR

Several schemes are considered for comparison with the proposed SMMSE decoder. As mentioned before, all these schemes reconstruct the prediction residues $\hat{\mathbf{y}}_n$ (or select the corresponding index $\hat{i}_n$), which is then fed to an ordinary DPCM decoder. In our experiment set up, this can be written as

$$\hat{\mathbf{x}}_n = \mathbf{A}\hat{\mathbf{x}}_{n-1} + \hat{\mathbf{y}}_n. \tag{40}$$

As baselines for comparisons, we consider the maximum-likelihood (ML) decoder given by

$$\hat{i}_n = \arg\min_{I_n \in \mathcal{J}} P(J_n|I_n) \tag{41}$$

and the basic MMSE decoder given by

$$\hat{\mathbf{y}}_n = \sum_{I_n \in \mathcal{J}} E[\mathbf{Y}_n|I_n]P(J_n|I_n). \tag{42}$$

Both the ML decoder and the MMSE decoder do not use any of the available residual redundancy for reconstruction.

The sequence MAP (SMAP) decoder is also considered for reconstruction of predictively encoded signals [11], [13], [19], [30]. The SMAP decoder evaluated here exploits the residual redundancy in the source-coder output with a Markov model of order γ. It decodes the prediction residues corresponding to the most probable transmitted sequence of symbols using a Viterbi-style decoder. The reconstructed prediction residues are then fed to an ordinary DPCM decoder. The trellis is constructed as described in Section III, and the metric corresponding to branch (S_{k-1}, S_k) is given by $\log[P(J_k|I_k)P(I_k|S_{k-1})]$.

The maximal SNR (MSNR) receiver presented in [30] is also considered for comparison. The MSNR decoding rule is given by

$$\hat{i}_n = \arg\min_{i_n \in \mathcal{J}} \sum_{i_k \in \mathcal{J}} d(i_n, i_k)P(J_n|I_n = i_k) \times P(I_n = i_k|\underline{I}_{n-1}^{n-\gamma} = \underline{\hat{i}}_{n-1}^{n-\gamma}) \tag{43}$$

where

$$d(i_n, i_k) = (\tilde{\mathbf{Y}}(i_n) - \tilde{\mathbf{Y}}(i_k))(\tilde{\mathbf{Y}}(i_n) - \tilde{\mathbf{Y}}(i_k))' \tag{44}$$

and $\tilde{\mathbf{Y}}(i_n)$ denotes the residue codeword corresponding to the index i_n; here, we have $\tilde{\mathbf{Y}}(i_n) = E[\mathbf{Y}_n|I_n = i_n]$. The MSNR decoder suffers from error propagation, since it is designed based on the assumption of the correctness of the previously decoded signals. This is observed from Fig. 3, which shows that the performance of the MSNR decoder degrades with the increase of redundancy order γ at high channel-error rates. In fact, it appears that by increasing γ, the gain due to the extra use of the residual redundancy is removed by the loss due to the error propagation. Note that in our experiments, for the cases where two or more symbols produce the same value for the distortion function of (43), we adopted a rule to select the symbol with the highest probability $P(J_n|I_n)$. We found that a trivial selection among these codewords, especially at very low error rates, results in error propagation and degrades the performance.

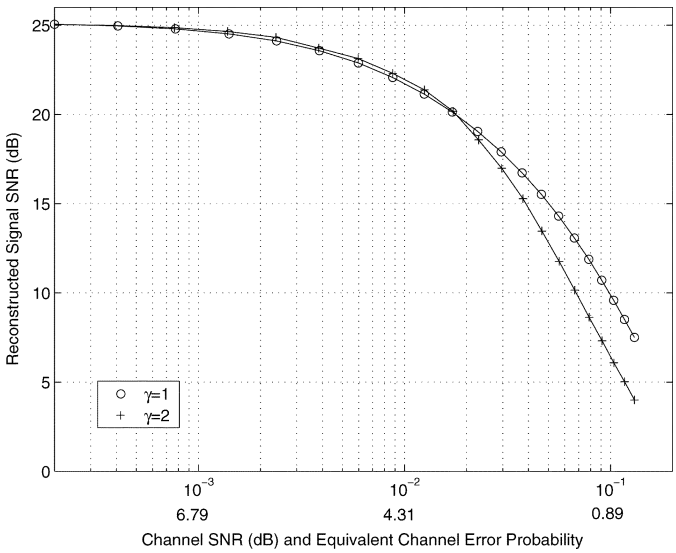

Fig. 3. Performance of the MSNR decoder for transmission of source A over an SOC, with binary index assignment, when different levels of residual redundancy are exploited at the decoder ($M = 8$).

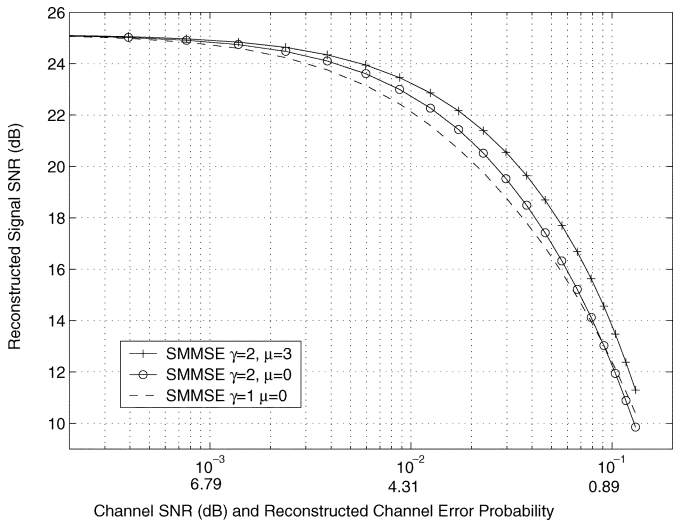

Fig. 4. Performance of the SMMSE decoder, effect of μ or the effective memory length of the decoder ($\gamma = 2$, $\delta = 0$) for source A. Binary index assignment is used.

B. SMMSE Decoder Numerical Results

Fig. 4 demonstrates the effect of the effective decoder memory length μ on the performance of the proposed SMMSE decoder for reconstruction of source A transmitted over a noisy channel. It is seen that increasing μ noticeably enhances the performance. For the case with $\mu = 3$, a gain of more than 1.5 dB in reconstructed signal SNR is achieved over the case with $\mu = 0$. Also depicted in Fig. 4 is the performance of the SMMSE decoder for the case with $\gamma = 1$, $\mu = 0$ for comparison. This is, in fact, equivalent to the sequence-based MMSE

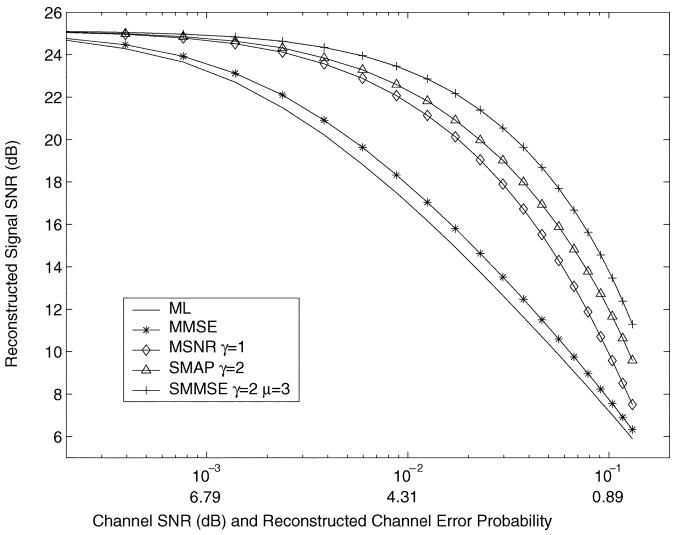

Fig. 5. Performance comparison of the SMMSE decoder with SMAP ($\delta = 0$), MSNR, basic MMSE, and ML decoders using binary index assignment (source A).

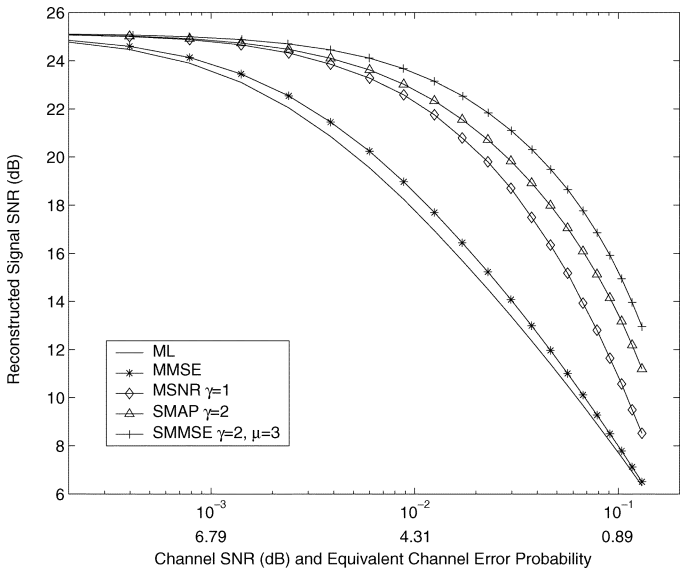

Fig. 6. Performance comparison of the SMMSE decoder with SMAP ($\delta = 0$), MSNR, basic MMSE, and ML decoders using Gray index assignment (source A).

decoder of [20], employed for the reconstruction of predictively encoded signals. Furthermore, since the case with $\delta = 0$ is considered, this also is equivalent to the instantaneous MMSE decoder of [19], used for decoding a DPCM signal. In comparison, the performance advantage of the proposed SMMSE decoder is noticeable, and reaches 1.85 dB at a channel SNR of 1.5 dB.

Fig. 5 provides a performance comparison between the proposed SMMSE decoder and the SMAP decoder. Also, the performance of the MSNR decoder of (43) for $\gamma = 1$ and the basic MMSE decoder of (42), as well as the ML decoder of (41), are depicted in the same figure. These results are obtained using source A and binary index assignment. It is seen that the proposed SMMSE decoder provides an effective solution for the reconstruction of predictive coded signals transmitted over a noisy channel. It outperforms the SMAP decoder by nearly 2 dB. The SMMSE decoder also gains as high as 8 dB compared with the ML decoder, and as high as 4 dB compared with the MSNR decoder.

Similar performance comparisons when Gray coding is used for index assignment are presented in Fig. 6. Noticeable gains are obtained using the Gray coding scheme, however, different schemes still compare the same, with the proposed SMMSE decoder providing the best performance. The gains achieved by employing the Gray coding as opposed to binary index assignment for ML, MMSE, MSNR, SMAP, and SMMSE decoders reach 0.77, 0.64, 1.01, 1.59, and 1.66 dB, respectively.

As mentioned, the simulation results are obtained by transmitting a test-data set over ten realizations of the noisy channel. The curves of Figs. 4–6 depict the average performance values. Table III provides the peak deviation of reconstructed SNR with respect to the average value for different methods. This is a measure of the variability of the results, with respect to different random seeds used for noisy channel realizations. Considering the large test-data set used, only very small sensitivity is observed.

We also evaluated the performance of various source decoders for reconstruction of source B transmitted over a noisy channel. These results are presented in Table IV. As expected, increasing residual redundancy order γ, and/or sequence length μ, constantly improves the performance. At $\gamma = 2$, $\mu = 3$, the proposed SMMSE decoder outperforms the basic MMSE decoder by more than 1.75 dB, and the decoder of [19] and [20] employed for reconstruction of predictively encoded signals, corresponding to the case with $\mu = 0$, $\gamma = 1$, by more than 0.9 dB. These gains are achieved by exploiting the small amount of residual redundancy available at the output of the DPCM source coder for source B, as given in Table II.

VI. Conclusions

In this paper, the problem of reconstruction of predictively encoded signals over noisy channels is considered. Due to suboptimality of the source coder, there is a residual redundancy in its output stream which is modeled by a γ-order Markov model. We present an SMMSE decoder which is formulated to minimize the MSE in the reconstruction of original signal (input to the source coder) at the receiver. This is different from the previous approaches, which aim at decoding the data samples received over the channel (prediction residues, output of the source coder). Numerical results are presented which demonstrate the effectiveness of the proposed algorithm. The complexity of the proposed technique grows with the increase of residual redundancy order γ and the effective memory length μ (see Sections III-C.I and III-C.II for details). An interesting future research direction for MMSE source decoding is, therefore, to investigate methods of reducing the complexity of these algorithms. Recent such developments for the case of memoryless VQ can be found in [22] and [36]. Other ideas for future research in this direction are suggested by the reviewers of this paper as follows. First, the proposed SMMSE decoder can be adapted for

TABLE III
PEAK DEVIATION FROM AVERAGE PERFORMANCE (%) FOR DECODING THE TEST DATABASE (SOURCE A) OVER TEN DIFFERENT REALIZATIONS OF THE NOISY CHANNEL USING BINARY INDEX ASSIGNMENT

Method	ML	MMSE	MSNR $\gamma = 1$	SMAP $\gamma = 2$	SMMSE $\gamma = 2, \mu = 3$
Peak Deviation (%)	0.57	0.36	0.86	1.24	0.24

TABLE IV
RECONSTRUCTED SIGNAL SNR (DB) FOR TRANSMISSION OF FIRST-ORDER GAUSS–MARKOV SOURCE OVER A NOISY CHANNEL ($\delta = 0$). DIFFERENT SMMSE DECODERS ARE IDENTIFIED BY PARAMETERS γ, RESIDUAL REDUNDANCY ORDER, AND μ SEQUENCE LENGTH

Channel SNR (dB)	BER	Basic MMSE	$\gamma = 1$ $\mu = 0$	$\gamma = 2$ $\mu = 0$	$\gamma = 1$ $\mu = 3$	$\gamma = 2$ $\mu = 3$
-2.0	0.130	5.48	6.26	6.94	6.64	7.19
-1.0	0.104	6.56	7.46	8.10	7.83	8.34
0.0	0.079	7.84	8.83	9.43	9.19	9.65
1.0	0.056	9.35	10.41	10.94	10.73	11.15
2.0	0.037	11.06	12.21	12.67	12.48	12.85
3.0	0.023	13.07	14.23	14.60	14.44	14.75
4.0	0.012	15.33	16.40	16.70	16.56	16.82
5.0	0.0059	17.71	18.58	18.80	18.69	18.88
6.0	0.0024	19.89	20.44	20.57	20.50	20.61
7.0	0.00075	21.39	21.62	21.67	21.65	21.69
8.0	0.00018	22.06	22.13	22.14	22.13	22.14

decoding variable-rate source codes; and second, the same sequence-decoding feature of this decoder can be implemented in a SMAP framework with a Viterbi-style search. The latter will lead to a reduced complexity at the expected cost of a certain level of performance degradation.

APPENDIX

The APPs of states within the source trellis diagram, described in Section III, are calculated by the following forward–backward equation:

$$\begin{aligned} P(S_n|\underline{J}_{n+\delta}) &= C_1 \cdot P(S_n, \underline{J}_n, \underline{J}_{n+\delta}^{n+1}) \\ &= C \cdot P(S_n|\underline{J}_n) \cdot P(\underline{J}_{n+\delta}^{n+1}|S_n, \underline{J}_n) \\ &= C \cdot P(S_n|\underline{J}_n) \cdot P(\underline{J}_{n+\delta}^{n+1}|S_n) \end{aligned} \quad (45)$$

where $C_1 = 1/P(\underline{J}_{n+\delta})$ and $C = P(\underline{J}_n)/P(\underline{J}_{n+\delta})$. The forward term is given by

$$\begin{aligned} &P(S_n|\underline{J}_n) \\ &\quad = C_1 \cdot P(S_n, \underline{J}_n) \\ &\quad = C_1 \cdot P(\underline{J}_{n-1}) \cdot P(S_n|\underline{J}_{n-1}) \cdot P(J_n|S_n, \underline{J}_{n-1}) \\ &\quad = C \cdot P(J_n|I_n) \cdot P(S_n|\underline{J}_{n-1}) \\ &\quad = C \cdot P(J_n|I_n) \cdot \sum_{S_{n-1}} P(S_n|S_{n-1}, \underline{J}_{n-1}) \cdot P(S_{n-1}|\underline{J}_{n-1}) \\ &\quad = C \cdot P(J_n|I_n) \cdot \sum_{S_{n-1}} P(S_n|S_{n-1}) \cdot P(S_{n-1}|\underline{J}_{n-1}) \end{aligned} \quad (46)$$

in which $C_1 = 1/P(\underline{J}_n)$ and $C = P(\underline{J}_{n-1})/P(\underline{J}_n)$. The backward term is calculated as shown in the first equation at the bottom of the page. Using (3) and (5), we have $P(\underline{J}_{n+\delta}^{n+1}|S_n, \underline{I}_{n+\delta}^{n+1}) = \prod_{k=1}^{\delta} P(J_{n+k}|I_{n+k})$, which simplifies the backward equation to (47), shown at the bottom of the page, which is calculated recursively at each time instant, starting from

$$P(J_{n+\delta}|S_{n+\delta-1}) = \sum_{I_{n+\delta}} P(J_{n+\delta}|I_{n+\delta}) P(I_{n+\delta}|S_{n+\delta-1})$$

and continuing backward in δ steps until $P(\underline{J}_{n+\delta}^{n+1}|S_n)$ is found. In summary, the recursions are composed of three parts: 1) the reliability of received data in the previous (next) time instants, i.e., $P(S_{n-1}|\underline{J}_{n-1})$ in (46) [$P(\underline{J}_{n+\delta}^{n+2}|S_{n+1})$ in (47)]; 2) the *a priori* information of the source, or the residual redundancy $P(I_n|S_n)$; and 3) the channel-related symbol probabilities $P(J_n|I_n)$. The second part is stored and available at the receiver; the third part is derived from the received data from the channel; and the first part is either recursively computed or is available from previous computations at the receiver.

$$\begin{aligned} P(\underline{J}_{n+\delta}^{n+1}|S_n) &= \sum_{I_{n+1}} P(\underline{J}_{n+\delta}^{n+1}|I_{n+1}, S_n) \cdot P(I_{n+1}|S_n) \\ &= \sum_{I_{n+1}} P(I_{n+1}|S_n) \sum_{I_{n+2}} P(\underline{J}_{n+\delta}^{n+1}|S_n, I_{n+1}, I_{n+2}) \cdot P(I_{n+2}|S_{n+1}) \\ &= \sum_{I_{n+1}} P(I_{n+1}|S_n) \sum_{I_{n+2}} P(I_{n+2}|S_{n+1}) \cdots \sum_{I_{n+\delta}} P(I_{n+\delta}|S_{n+\delta+1}) \cdot P(\underline{J}_{n+\delta}^{n+1}|S_n, \underline{I}_{n+\delta}^{n+1}) \end{aligned}$$

$$\begin{aligned} P(\underline{J}_{n+\delta}^{n+1}|S_n) &= \sum_{I_{n+1}} P(J_{n+1}|I_{n+1}) P(I_{n+1}|S_n) \sum_{I_{n+2}} P(J_{n+2}|I_{n+2}) P(I_{n+2}|S_{n+1}) \cdots \sum_{I_{n+\delta}} P(J_{n+\delta}|I_{n+\delta}) P(I_{n+\delta}|S_{n+\delta-1}) \\ &= \sum_{I_{n+1}} P(J_{n+1}|I_{n+1}) P(I_{n+1}|S_n) P(\underline{J}_{n+\delta}^{n+2}|S_{n+1}) \end{aligned} \quad (47)$$

ACKNOWLEDGMENT

The authors wish to thank Dr. D. J. Miller and Dr. K. Sayood for some insightful discussions. Also, we thank the anonymous reviewers, whose constructive comments substantially improved this paper.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
[2] M. Bystrom and J. W. Modestino, "Combined source-channel coding schemes for video transmission over an additive white Gaussian noise channel," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 880–890, June 2000.
[3] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1412–1424, Sept. 1997.
[4] S. Gadkari and K. Rose, "Unequally protected multistage vector quantization for time-varying CDMA channels," *IEEE Trans. Commun.*, vol. 49, pp. 1045–1054, June 2001.
[5] E. Ayanoglu and R. M. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 855–865, Nov. 1987.
[6] H. Jafarkhani and N. Farvardin, "Design of channel-optimized vector quantizers in the presence of channel mismatch," *IEEE Trans. Commun.*, vol. 48, pp. 118–124, Jan. 2000.
[7] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.
[8] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Theory for transmission of vector quantization data," in *Speech Coding and Synthesis*, W. Kleijn and K. Paliwal, Eds. New York: Elsevier, 1995, pp. 347–396.
[9] ——, "Vector quantization for speech transmission," in *Speech Coding and Synthesis*, W. Kleijn and K. Paliwal, Eds. New York: Elsevier, 1995, pp. 311–345.
[10] K. Sayood, H. H. Otu, and N. Demir, "Joint source/channel coding for variable length codes," *IEEE Trans. Commun.*, vol. 48, pp. 787–794, May 2000.
[11] F. Lahouti and A. K. Khandani, "Efficient source decoding over memoryless noisy channels using higher-order Markov models," *IEEE Trans. Inform. Theory*, to be published.
[12] F. Lahouti, "Quantization and reconstruction of sources with memory," Ph.D. dissertation, Univ. Waterloo, Dept. E&CE, Waterloo, ON, Canada, 2002.
[13] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Trans. Commun.*, vol. 39, pp. 838–845, June 1991.
[14] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, pp. 2449–2457, Sept. 1995.
[15] K. Sayood, L. Fuling, and J. D. Gibson, "A constrained joint source/channel coder design," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1584–1593, Dec. 1994.
[16] F. I. Alajaji, N. Phamdo, and T. E. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 325–336, Sept. 1996.
[17] A. Ruscitto and E. M. Biglieri, "Joint source and channel coding using turbo codes over rings," *IEEE Trans. Commun.*, vol. 46, pp. 981–984, Aug. 1998.
[18] T. Fazel and T. E. Fuja, "Joint source-channel decoding of block-encoded compressed speech," in *Proc. Conf. Information Sciences and Systems*, Mar. 2000, pp. FA5-1–FA5-6.
[19] N. Phamdo and N. Farvardin, "Optimal detection of discrete Markov sources over discrete memoryless channels—Applications to combined-source channel coding," *IEEE Trans. Inform. Theory*, vol. 40, pp. 186–103, Jan. 1994.
[20] D. J. Miller and M. Park, "A sequence-based approximate MMSE decoder for source coding over noisy channels using discrete hidden Markov models," *IEEE Trans. Commun.*, vol. 46, pp. 222–231, Feb. 1998.
[21] T. Fingscheidt and P. Vary, "Softbit speech decoding: A new approach to error concealment," *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 240–251, Mar. 2001.
[22] F. Lahouti and A. K. Khandani, "Approximating and exploiting the residual redundancies— Applications to efficient reconstruction of speech over noisy channels," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 2, Salt Lake City, UT, May 2001, pp. 721–724.
[23] M. Adrat, J. Spittka, S. Heinen, and P. Vary, "Error concealment by near-optimum MMSE estimation of source codec parameters," in *Proc. IEEE Workshop on Speech Coding*, 2000, pp. 84–86.
[24] F. Lahouti and A. K. Khandani, "An efficient MMSE source decoder for noisy channels," in *Proc. Int. Symp. Telecommunications*, Tehran, Iran, Sept. 2001, pp. 784–787.
[25] T. Fingscheidt, T. Hindelang, R. V. Cox, and N. Seshadri, "Joint source-channel (de-)coding for mobile communications," *IEEE Trans. Commun.*, vol. 50, pp. 200–212, Feb. 2002.
[26] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1662–1670, Sept. 2001.
[27] M. Adrat, P. Vary, and J. Spittka, "Iterative source-channel decoder using extrinsic information from softbit-source decoding," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 4, Salt Lake City, UT, May 2001, pp. 2653–2656.
[28] K. Sayood, *Introduction to Data Compression*. San Francisco, CA: Morgan Kaufmann, 2000.
[29] *Digital Coding of Waveforms*, Bell Tel. Labs. Inc., 1984.
[30] S. Emami and S. L. Miller, "DPCM picture transmission over noisy channels with the aid of a Markov model," *IEEE Trans. Image Processing*, vol. 4, pp. 1473–1481, Nov. 1995.
[31] M. Park and D. J. Miller, "Improved image decoding over noisy channels using MMSE estimation and a Markov mesh," *IEEE Trans. Image Processing*, vol. 8, pp. 863–867, June 1999.
[32] R. Link and S. Kallel, "Optimal use of Markov models for DPCM picture transmission over noisy channels," *IEEE Trans. Image Processing*, vol. 48, pp. 1702–1711, Oct. 2000.
[33] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.
[34] K. Y. Chang and R. W. Donaldson, "Analysis, optimization, and sensitivity study of differential PCM systems operating on noisy communication channels," *IEEE Trans. Commun.*, vol. COM-20, pp. 338–350, June 1972.
[35] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.
[36] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1293–1307, May 1999.
[37] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

Farshad Lahouti (M'02) received the B.Sc. degree from the University of Tehran, Tehran, Iran in 1997, and the Ph.D. degree in 2002 from the University of Waterloo, Waterloo, ON, Canada, both in electrical and computer engineering.

He is currently a Research Assistant Professor with the University of Waterloo. Part of his work is the main component of a research project [co-funded by Nortel Networks and Communications and Information Technology (CITO)] that has been documented as a CITO success story. His research interests include source/channel coding for multimedia communication over wireless/packet networks and DSP applications in digital communications.

Dr. Lahouti is the recipient of the Natural Sciences and Engineering Research Council of Canada award, as well as seven other provincial and departmental awards.

Amir K. Khandani (M'02) received the M.A.Sc. degree from University of Tehran, Tehran, Iran, and the Ph.D. degree from McGill University, Montreal, QC, Canada, in 1985 and 1992, respectively.

Following graduation, he worked for one year as a Research Associate at INRS-Telecommunication, Montreal. In 1993, he joined the Department of Electrical and Computer Engineering of the University of Waterloo, Waterloo, ON, Canada, where he is currently a Professor and a Senior Industrial Research Chair holder funded by Nortel Networks and the Natural Sciences and Engineering Research Council of Canada (NSERC).

Dr. Khandani is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of Coding and Communication Theory.