Efficient MMSE Source Decoding Over Noisy Channels

Farshad Lahouti and Amir K. Khandani

Technical Report UW-E&CE#2002-3

Department of Electrical & Computer Engineering

University of Waterloo

Waterloo, Ontario, Canada, N2L 3G1

Feb. 19, 2002

Efficient MMSE Source Decoding Over Noisy Channels

Farshad Lahouti, Amir K. Khandani February 19, 2002

Abstract

Exploiting the residual redundancy in a source coder output stream during the decoding process has been proven to be a bandwidth efficient way to combat the noisy channel degradations. Researchers have recently developed techniques to employ this redundancy to either assist the channel decoder for improved performance or design better source decoders. However, the method used for modeling the redundancy is a first-order Markov model which fails to encapsulate all the remaining redundancies. In this work, we present a family of solutions for the asymptotically optimum Minimum Mean Squared Error reconstruction of a source over memoryless noisy channels when the redundancy in the source encoder output stream is exploited in the form of a γ-order Markov model ($\gamma \geq 1$) and a delay of δ, $\delta > 0$, is allowed in the decoding process. We demonstrate that the proposed solutions provide a wealth of trade-offs between computational complexity and the memory requirements. We also present a simplified MMSE decoder which is optimized to minimize the computational complexity. Considering the same problem setup, we present several other Maximum A Posteriori symbol and sequence decoders as well. Numerical results are presented which demonstrate the efficiency of the proposed algorithms.

Keywords

Joint source channel coding, residual redundancies, source decoding, MMSE estimation, MAP detection, forward backward recursion, Markov sources

This work is funded in part by the Natural Sciences and Engineering Research Council of Canada. This work has been presented in part at the International Symposium on Telecommunications, Tehran, Iran, 2001. The authors are affiliated with the Coding & Signal Transmission Lab., Dept. of E&CE, University of Waterloo, Waterloo, ON, N2L 3G1, Canada, Email: (farshad, khandani)@cst.uwaterloo.ca.

I. Introduction

An important result of the Shannon's celebrated paper [1], is that the source and channel coding operations can be separated without any loss of optimality. This has been the basic idea of enormous research endeavors in separate treatment of source and channel coders. However, in practise, due to strict design constraints such as limited transmission bandwidth, high error protection requirements along with restricted delay and limitations on the complexity of the systems involved, the joint design of source and channel coders has found increasing interest.

Researchers have taken several paths toward the joint design of source and channel coders. A class of joint source channel coders are designed by attempting to optimally allocate a fixed bitrate between the source and channel coders to achieve the maximum overall system performance. Examples of the works in this class are present in [2]-[6]. The applications span across the areas of speech, image and video coding such as that of Modestino *et al.* on image coding using the discrete cosine transform with convolutional channel coding [2], the work of Moore and Gibson on DPCM speech coding with self orthogonal convolutional coding [3] and the work of Bystrom and Modestino on combined source channel coding for video transmission [4].

Other methods of joint source and channel coding include the systems designed based on *Unequal Error Protection*, *Optimized Index Assignment*, *Channel Optimized Quantization* and more recently *exploiting the source residual redundancies*. In some applications a combination of these techniques is employed for a greater protection over noisy channels.

Systems designed with Unequal Error Protection provide better error protection for the parts of the source coder output stream which have a greater contribution in the objective or subjective quality of the reconstructed source. One good example of this technique is the North American IS-641 [7] standard which accommodates three different classes of error protection for different output bits of a CELP-based speech coder. A related classical work is the work of Sundberg [8] in which he analyzed the effect of error in different bits on the reconstruction of a PCM coded signal. Examples of more recent applications of UEP is present in [9] and [10].

The Index Assignment technique provides more robustness to channel errors by assigning the quantizer outputs to encoder indices in a way that possible bit errors create a lower level of distortion in the reconstructed data. One usual advantage of the index assignment is that it does not degrade the performance during the clean channel conditions. For a review of different index assignment techniques refer to [11].

In Channel Optimized Quantization, the quantization levels are designed to optimize the performance of the system in the presence of channel noise. Two classic works in this area are those of Kurtenbach and Wintz [12] on scalar quantization over noisy channels and Chang and Donaldson [13] on the design of a DPCM system for transmission over a discrete memoryless channel. Other works on channel optimized quantization include the works of Kumazawa *et. al.* [14] and Farvardin and Vaishampayan [15] on vector quantization over noisy channels as well as the works of Dunham and Gray [16] and Ayanoglu and Gray [17] on joint source channel trellis coding. Examples of more recent works in this class are present in [18]-[20]. For a more comprehensive review of the techniques for channel optimized quantization, the interested reader is referred to [11],[20],[21].

More recently in this venue, exploiting the *residual redundancy* [22] in the output of the source coders for improved reconstruction over noisy channels has found increasing attention [22]-[41]. This redundancy is due to the suboptimal source coding which is caused by, e.g., a constraint on complexity or delay. Researchers have used the residual redundancy for enhanced channel decoding, e.g., [23]-[27] or for effective source decoding, e.g., [28]-[32]. The problem is formulated in the form of a *Maximum A Posteriori* detection or a *Minimum Mean Squared Error* estimation problem. In [33], Phamdo and Farvardin proposed instantaneous MAP and MMSE decoders as well as a MAP sequence decoder using the residual redundancies exploited by a first-order Markov model. Later in [34], a sequence-based MMSE decoder was suggested that benefits from the redundancies of both the past and future samples. Source decoding over channels with memory using the residual redundancies has been considered in [35] and [36].

In [37], it was demonstrated that the use of residual redundancies both at the source and channel decoder could lead to improved performance. In the same direction, iterative source and channel decoding schemes were presented in [38]-[40]. The effectiveness of these techniques have lead the researchers to new horizons. In [41], it is suggested that intentional leaving of the redundancy of the source, through the use of simpler source coders, could result in higher performance when this redundancy is exploited effectively at the decoder. This higher performance is either attributed to lower overall system complexity or better trade-offs of bandwidth between the source and channel coding.

A. Contributions of the manuscript

The recent literature clearly demonstrate the benefit of exploiting the residual redundancies in reconstructing the data received over noisy channels. However, it has limited itself to modeling the redundancy with a first-order Markov model, which does not necessarily encapsulate all the remaining redundancy. In this work, we present a family of solutions for the asymptotically optimum MMSE reconstruction of a source over memoryless noisy channels when the redundancy in the source encoder output stream is exploited in the form of a γ-order Markov model ($\gamma \geq 1$) and a delay of δ, $\delta > 0$, is allowed in the decoding process. We demonstrate that the proposed solutions provide a wealth of trade-offs between computational complexity and the memory requirements. We also present a simplified MMSE decoder which is optimized to minimize the computational complexity. Considering the same problem setup, we present several other Maximum A Posteriori symbol and sequence decoders as well. Finally, we study the effect of different system parameters and characteristics on the performance of the proposed decoders.

The organization of this paper is as follows. In section II, an overview of the system and the channel model used is described. In section III, the MMSE decoding problem statement and solutions are presented. In section IV, the MAP decoding problem statement and solutions are presented. Section V, includes the numerical results and various comparisons which demonstrate the effectiveness of the proposed schemes. We conclude this article in section VI.

II. Preliminaries

A. Notations

The notations used in this article 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 a confusion. The vectors are shown bold faced, 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 representing an integer value, 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 $\mathcal{J}^N$ that consists of N dimensional vectors whose components are taken from $\mathcal{J}$.

B. System Overview

The block diagram of the system is shown in Figure 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: the quantizer Q and the index generator $\mathcal{I}$. The quantizer maps the input sample $\mathbf{X} \in \mathcal{R}^N$ to one of the reconstruction points or *codewords* in the codebook $\mathcal{C}$, $\mathcal{C} \subset \mathcal{R}^N$. The index generator then maps this codeword to the an *index (symbol)* I in the index set $\mathcal{J}$. The bitrate of the quantizer r is given by $\lceil \log_2 M \rceil$ bits/symbol (or $\lceil \log_2 M \rceil / N$ bits/dim). We assume the source encoder is memoryless, i.e., the mapping of $\mathbf{X}_n$ to I_n is independent from the past and future values of the encoder input and output.

Fig. 1. Overview of the system

At the receiver, for each transmitted r-bit index $I = i$, a vector J with r components is received which provides information about I. The reconstructor (source decoder) maps J to an output sample $\hat{\mathbf{X}}$. In this reconstruction, the source decoder may use the previously received signals or some of the future samples as well.

C. Channel Model

The channels considered in this work are described by a pdf $P(J_n|I_n)$. We assume that the channel is memoryless without intersymbol interference 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.

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

This results in the followings,

$$P(J_n = j_n | \underline{I}_n = \underline{i}_n) = P(J_n = j_n | I_n = i_n), \tag{2}$$

$$P(\underline{J}_n = \underline{j}_n | \underline{I}_n = \underline{i}_n) = \prod_{k=1}^{n} P(J_k = j_k | I_k = i_k). \tag{3}$$

An example is a BPSK modulation over a channel with AWGN which produces soft outputs as,

$$j_n^{(m)} = s(i_n^{(m)}) + \eta_n^{(m)}, \quad m = 1, \ldots, r. \tag{4}$$

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 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,

$$P(J_n = j_n | I_n = i_n) = \prod_{m=1}^{r} P(j_n^{(m)} | i_n^{(m)}). \tag{5}$$

In this work, we refer to such a channel as the Soft Output Channel model. The Binary Symmetric Channel model is also based on the same Equation (4), when a hard decision is made on the received soft outputs. If the resulting bit error probability is denoted by ϵ, then the relationship between the transmitted and the received symbols is given by,

$$P(J_n = j_n | I_n = i_n) = (\epsilon)^{h(i_n, j_n)} (1 - \epsilon)^{r - h(i_n, j_n)}, \tag{6}$$

where j_n is the received binary codeword in $\mathcal{J}$ and $h(i_n, j_n)$ is the Hamming distance between indices i_n and j_n. In the followings, 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 Equation (1) is valid.

III. MMSE Decoding: Problem Statement and Solutions

Consider the case where due to the suboptimality of the source coder, there is a residual redundancy in its output stream. This redundancy is in the form of a non-uniform distribution or a memory in the sequence of the transmitted symbols. Our objective is to design an effective reconstructor (source decoder) which exploits this redundancy and produces the Minimum Mean Squared Error estimate of the source sample $\mathbf{X}_n$ given the received sequence $\underline{J}_{n+\delta} = \underline{j}_{n+\delta} = [j_1, j_2, \ldots, j_{n+\delta}]$. Based on the fundamental theorem of Estimation Theory, this is given by,

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n | \underline{J}_{n+\delta} = \underline{j}_{n+\delta}] \tag{7}$$

which minimizes the expected squared error of estimation,

$$E[(\mathbf{X}_n - \hat{\mathbf{X}}_n)'(\mathbf{X}_n - \hat{\mathbf{X}}_n)] \tag{8}$$

In Equation (7), we have,

$$E[\mathbf{X}_n | \underline{J}_{n+\delta}] = \sum_{\underline{I}_{n+\delta} \in \mathcal{J}^{n+\delta}} E[\mathbf{X}_n | \underline{J}_{n+\delta}, \underline{I}_{n+\delta}] P(\underline{I}_{n+\delta} | \underline{J}_{n+\delta}) \tag{9}$$

and noting that condition on $\underline{I}_{n+\delta} = \underline{i}_{n+\delta}$, $\mathbf{X}_n$ is independent of $\underline{J}_{n+\delta}$, we have,

$$E[\mathbf{X}_n|\underline{J}_{n+\delta}, \underline{I}_{n+\delta}] = E[\mathbf{X}_n|\underline{I}_{n+\delta}]$$

which forms the *decoder codebook*. Therefore, the optimal decoder at time n requires a sum over $M^{n+\delta}$ elements of the decoder codebook. It is seen that in this case both computational complexity and the memory requirement grow exponentially with time, leading to an impractical scheme. In the next subsection, we develop an asymptotically optimum MMSE decoder for the cases where the residual redundancy is modeled by a γ-order Markov model and show that it leads to a feasible decoder. Subsequently in section III-B, we present a simplified MMSE decoder.

A. An Asymptotically Optimum MMSE Decoder

Assuming that the source $\mathbf{X}$ has a memory that asymptotically decays with time, for sufficiently large values of τ, $\tau \in \mathcal{Z}$, we have,

$$E[\mathbf{X}_n|\underline{I}_{n+\delta}] \approx E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-\tau}]. \tag{10}$$

This simplifies the optimum decoder to the following decoder,

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

which is asymptotically optimal for $\tau \gg 0$. We refer to the decoder of Equation (11) as the Asymptotically Optimum Minimum Mean Squared Error (AOMMSE) decoder. It describes the decoded signal in the form of the weighted average of the codewords $E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-\tau}]$. Note that $E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-\tau}]$ provides a finer reconstruction of the source symbols than the codewords $E[\mathbf{X}_n|I_n]$ used at the transmitter side. Particularly, as seen in the results section, if the channel is noise free, the signal decoded at the receiver will be closer to the original source (higher signal to noise ratio) when compared to the source encoded signal at the transmitter. The extra information is provided by the memory between the symbols $[I_{n-\tau}, \ldots, I_{n+\delta}]$.

At each time instant, the decoder needs to calculate the instantaneous values of the weights in Equation (11) or the probabilities $P(\underline{I}_{n+\delta}^{n-\tau}|\underline{J}_{n+\delta})$. To calculate these probabilities, we assume that due to the residual redundancy at the encoder output, the encoder output symbols form a γ-order, $\gamma > 0$, Markov model. In the following sections, we first present a solution to compute these probabilities and then consider alternative solutions to optimize the computational complexity. These solutions are valid for all values of $\tau \geq \gamma$ (the case $\tau < \gamma$ will be straight forward).

A.1 A basic solution

To compute the a posteriori probabilities required in the AOMMSE decoder of Equation (11), we use a trellis structure. The trellis structure models the symbols I_n and their assumed γ-order Markov property due to the residual redundancies. In this section, we first describe the trellis structure and subsequently, present a solution based on this structure.

We consider a trellis structure in which, the states at time n correspond to the ordered set,

$$S_n = (I_{n-\gamma+1}, I_{n-\gamma+2}, \ldots, I_{n-1}, I_n), \qquad I_{n-k} \in \mathcal{J}, 0 \leq k < \gamma \tag{12}$$

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. The states now form a first-order Markov sequence, i.e., for states $S_1, S_2, \ldots, S_n$ that form an arbitrary path of the trellis, we have,

$$P(S_n|S_{n-1}, S_{n-2}, \ldots, S_1) = P(S_n|S_{n-1}). \tag{13}$$

Using this property and the memoryless assumption of the channel (see Equations (1)-(3)), in line with the BCJR algorithm [42], the probability of a particular state $S_n = s_n$ given the observed sequence $\underline{J}_n$ is calculated recursively as follows,

$$P(S_n|\underline{J}_n) = C.P(J_n|I_n) \,. \sum_{S_{n-1} \to S_n} P(I_n|S_{n-1}) P(S_{n-1}|\underline{J}_{n-1}) \tag{14}$$

where the summation is over a subset of M states in time step $n-1$ which are connected to the state S_n and C is a factor which normalizes the sum of the probabilities to one. The details of the derivation of Equation (14), which is referred to as the *forward* recursive equation, are provided in the Appendix-A. We use the notation C as the normalizing factor throughout this paper.

In the same direction, the probabilities of states given the observed sequence $\underline{J}_{n+\delta}$, $\delta \geq 0$ are calculated by the following equation,

$$P(S_n|\underline{J}_{n+\delta}) = C \,.P(S_n|\underline{J}_n) \,. P(\underline{J}_{n+\delta}^{n+1}|S_n) \tag{15}$$

where $\underline{J}_{n+\delta}^{n+1} = [J_{n+1}, J_{n+2}, \ldots, J_{n+\delta}]$. The Equation (15) is referred to as the *forward backward* equation in which the first term is the forward equation given in (14) and the second term is referred to as the *backward* equation and can be calculated recursively as follows,

$$P(\underline{J}_{n+\delta}^{n+1}|S_n) = \sum_{I_{n+1}\in\mathcal{J}} P(J_{n+1}|I_{n+1})\,.\,P(I_{n+1}|S_n)\,.\,P(\underline{J}_{n+\delta}^{n+2}|S_{n+1}) \tag{16}$$

where the recursion starts from,

$$P(J_{n+\delta}|S_{n+\delta-1}) = \sum_{I_{n+\delta}\in\mathcal{J}} P(J_{n+\delta}|I_{n+\delta})\,.\,P(I_{n+\delta}|S_{n+\delta-1}) \tag{17}$$

and continues backward in each time step. The details of the derivation of these equations are provided in the Appendix-A. The presented trellis structure and either of the forward and backward equations are used in the following sections for calculation of different symbol or sequence probabilities. We note that in each time step, the forward recursion of Equation (14) proceeds one step forward through the trellis while the backward term is recomputed over the entire backward window as indicated in Equations (16) and (17).

Now, using the presented trellis structure and the forward equation (14), the probabilities required for the asymptotically optimum MMSE decoding of Equation (11) are calculated recursively by the following ($\tau \geq \gamma > 0$, see Appendix-B for proof),

$$P(\underline{I}_{n+\delta}^{n-\tau}|\underline{J}_{n+\delta}) = C\,.\left[\prod_{k=-(\tau-\gamma)}^{\delta} P(J_{n+k}|I_{n+k})P(I_{n+k}|S_{n+k-1})\right] P(S_{n-(\tau-\gamma)-1}|\underline{J}_{n-(\tau-\gamma)-1}). \tag{18}$$

At each time instant the M^γ probabilities $P(S_n|\underline{J}_n)$ corresponding to each state are stored to be used in Equation (18) at the next time instant. In addition to the computations required to do this task, the complexity of the AOMMSE decoder is comprised of the cost to perform the multiplications and normalization in Equation (18), as well as the weighted average of the reconstruction rule in Equation (11). Clearly, these computations are not trivial and therefore, efficient alternative solutions are of particular interest. In the followings, we present efficient alternative exact solutions based on construction of an *extended trellis structure* of the source.

A.2 Solutions based on the extended trellis structure

The general problem of calculating the weights required in Equation (11), can be viewed as finding the probability of a sequence of states (symbols) within the structure of the source trellis given the entire history of the received signals or equivalently,

$$P(S_{n-L}, \ldots, S_{n-1}, S_n|\underline{J}_n) = P(\underline{I}_n^{n-(\gamma+L)+1}|\underline{J}_n), \quad L > 0,$$

where $L+1$ is the length of the sequence of states. The alternative solutions are provided based on different constructions of an extended trellis structure for the source coder output symbols. The states in this structure are referred to as the *super states* and are defined as,

$$SS_n = (I_{n-(\gamma+L')+1}, \ldots, I_{n-1}, I_n), \qquad I_{n-k} \in \mathcal{J}, 0 \leq k < \gamma + L', \tag{19}$$

where L', $0 \leq L' < L$ is referred to as the *(state set) extension factor* and is the number of symbols by which the states (Equation (12)) have been extended to form the super states. Similar to the original trellis, there are M branches leaving a (super) state in the extended trellis, where each branch corresponds to one symbol $I_n = i_n$, $i_n \in \mathcal{J}$. Therefore, each stage of the extended trellis still corresponds to one time step.

Based on the proposed extended trellis, a family of solutions to calculate the required a posteriori probabilities are given as follows (see Appendix-B for proof),

$$P(\underline{I}_n^{n-(\gamma+L)+1}|\underline{J}_n) = C\,.\left[\prod_{k=0}^{L-L'-1} P(J_{n-k}|I_{n-k})P(I_{n-k}|SS_{n-k-1})\right].P(SS_{n-(L-L')}|\underline{J}_{n-(L-L')}) \tag{20}$$

where,

$$P(I_{n-k}|SS_{n-k-1}) = P(I_{n-k}|S_{n-k-1}) \tag{21}$$

and $P(SS_{n-(L-L')}|\underline{J}_{n-(L-L')})$ are $M^{\gamma+L'}$ probabilities of super states which are stored in each time step. This term is updated by,

$$P(SS_l|\underline{J}_l) = C\,.\sum_{SS_{l-1}\to SS_l} P(J_l|I_l)P(I_l|SS_{l-1})P(SS_{l-1}|\underline{J}_{l-1}). \tag{22}$$

in which $l = n-(L-L')+1$ and the terms within the summation is available during the process of calculating Equation (20). The direct implementation of Equation (20) leads to a computational complexity[1] CC of,

$$CC = 2(L-L'+1)M^{L+\gamma} + (M+2)M^{L'+\gamma} \tag{23}$$

in which the first term is the cost due to the required multiplication of the terms and the required normalization and the second term includes the cost due to updating and normalizing

[1]In this work, the computational complexity is measured in terms of the number of floating point operations. Each addition, multiplication or comparison is considered as one floating point operation (flop).

the probability of the super states according to Equation (22). The memory requirement includes the fixed amount of static memory (ROM) required to store $M^{\gamma+1}$ transition probabilities, as well as the dynamic memory (RAM) required for the operation of the algorithm, which is $O(M^{\gamma+L'})$ based on the number of super states. This indicates that the family of solutions of Equation (20) provide a wealth of trade-offs of computational complexity and the memory requirement. The increase of L', reduces the computational complexity at the cost of an increase in memory requirement. It is important to note that any increase of L', $0 \leq L' < L$, beyond $L-1$ would lead to a solution which is suboptimal both in terms of computational complexity and the memory requirement.

More trade-offs of computational complexity and memory are possible considering the fact that the structure of the extended trellis is still based on the redundancy of the source as indicated in Equation (21). For example using Equation (21), the multiplying terms of Equation (20), can be calculated once for the M^{γ} states S_{n-k-1} and stored as a $M^{\gamma} \times M^{L-L'}$ matrix to be appropriately multiplied by the probability of super states. This reduces the corresponding computations in Equation (23) from $2(L-L')M^{L+\gamma}$ to $2(L-L')M^{L-L'+\gamma}$.

B. A Simplified MMSE Decoder

As mentioned before, the asymptotically optimum MMSE decoder presented in the previous section utilizes a codebook given by $E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-\tau}]$ which provides a finer reconstruction of the source when compared to the encoder codebook $E[\mathbf{X}_n|I_n]$. An approximation of interest to the AOMMSE decoder is to consider a decoder codebook identical to that of the encoder. We refer to this decoder as the MMSE decoder. Simplifying the Equation (7), the MMSE decoder is given by,

$$\hat{\mathbf{x}}_n = \sum_{I_n \in \mathcal{J}} E[\mathbf{X}_n|I_n] P(I_n|\underline{J}_{n+\delta}) \tag{24}$$

which describes the MMSE estimate in terms of the weighted average of the encoder (LBG [43]) codewords. The weights are the probability of receiving the corresponding symbol given the received sequence $\underline{J}_{n+\delta}$. It is noteworthy that in the trivial case where there is no memory between the symbols I_n (corresponding to $\gamma = 0$), the Equation (24) collapses to the basic MMSE reconstruction rule,

$$\hat{\mathbf{x}}_n = \sum_{I_n \in \mathcal{J}} E[\mathbf{X}_n|I_n] P(I_n|J_n) \tag{25}$$

in which the probability $P(I_n|J_n) = C.P(I_n).P(J_n|I_n)$, $C = \frac{1}{P(J_n)}$ includes the residual redundancy in the form of the non-uniform symbol a priori probabilities.

In the followings, we first present an efficient solution to calculate the required probabilities in Equation (24) based on the presented original trellis structure ($\gamma > 0$). Next, we investigate possible alternative solutions based on the extended trellis structure.

B.1 A basic solution

The a posteriori probability of a symbol I_n given the received sequence $\underline{J}_{n+\delta}$ is calculated as follows. Assuming that the encoded sequence contain a residual redundancy in the form of a γ-order, $\gamma \geq 1$, Markov model, we use the probabilities of states in the original trellis structure as described in section III-A.1. In particular when no delay is allowed in the decoding process, $\delta = 0$, we have,

$$P(I_n|\underline{J}_n) = \sum_{I_{n-\gamma+1}} \dots \sum_{I_{n-2}} \sum_{I_{n-1}} P(S_n|\underline{J}_n). \tag{26}$$

The Equations (14) and (26) together with the reconstruction rule of Equation (24) provide the instaneous (no delay allowed, i.e., $\delta = 0$) MMSE decoding of the source samples given the history of the received channel outputs.

We observe that the required symbol a posteriori probabilities can be alternatively calculated using the a posteriori probabilities of *any of the states* S_{n+m} *as long as* S_{n+m} *includes* I_n*, i.e.,* $0 \leq m \leq \gamma - 1$. As presented below, this is of particular interest when a delay of $\delta > 0$ is allowed in the decoding process. In such cases, this flexibility can be used to optimize the solution in terms of the complexity. We have,

$$P(I_n|\underline{J}_{n+\delta}) = \dots \textstyle\sum_{I_{n+k}} \dots \; P(S_{n+m}|\underline{J}_{n+\delta}), \tag{27}$$
$$\forall m \in \mathcal{Z}, 0 \leq m \leq \gamma - 1$$
$$k = m - \gamma + 1, \dots, m, \qquad k \neq 0.$$

where the probabilities of states $P(S_{n+m}|\underline{J}_{n+\delta})$ as described in section III-A.1, are given by the following forward backward equation,

$$P(S_{n+m}|\underline{J}_{n+\delta}) = C\,.P(S_{n+m}|\underline{J}_{n+m})\,.\,P(\underline{J}_{n+\delta}^{n+m+1}|S_{n+m}) \tag{28}$$
$$0 \leq m \leq \delta$$

in which the forward and the backward terms are given in Equations (14) and (16). In Equation (28), if the number of computations required for the forward and backward recursions (Equations

(14) and (16)) per time step is denoted by CC_{fwd} and CC_{bwd}, respectively, we have,

$$CC_{fwd} = (2M+3)\,M^{\gamma} \tag{29}$$

$$CC_{bwd} = 3(\delta - m)\,M^{\gamma+1} \tag{30}$$

where $\delta - m$ is the number of backward recursions required per time step. The overall complexity of computing Equation (27) is then given by[2]

$$CC = CC_{fwd} + CC_{bwd} + 2M^{\gamma} + 2M \tag{31}$$

Noting that only CC_{bwd} depends on m in Equation (31), to minimize the overall computational burden, we solve the following for the optimum value of m,

$$\begin{aligned} &\text{Minimize} \quad CC_{bwd} = 3(\delta - m)\,.\,M^{\gamma+1} \\ &\text{subject to} \quad 0 \le m \le \gamma - 1; \;\; 0 \le m \le \delta \end{aligned} \tag{32}$$

case 1. $\delta < \gamma$ In the cases where the delay is smaller than the assumed residual redundancy order, we are able to choose $m = \delta$ and eliminate the backward term. The probabilities in Equation (24) are then given as follows,

$$\begin{aligned} P(I_n|\underline{J}_{n+\delta}) = \ldots \textstyle\sum_{I_{n+k}} \ldots \; P(S_{n+\delta}|\underline{J}_{n+\delta}), \\ k = \delta - \gamma + 1, \ldots, \delta, \quad k \neq 0. \end{aligned} \tag{33}$$

case 2. $\delta \ge \gamma$ Alternatively, when the delay is larger than the assumed redundancy order, the CC_{bwd} is minimized when $m = \gamma - 1$, i.e., $\delta - \gamma + 1$ backward recursions is required. The probabilities in Equation (24) are now given by,

$$P(I_n|\underline{J}_{n+\delta}) = \sum_{I_{n+1}} \sum_{I_{n+2}} \cdots \sum_{I_{n+\gamma-1}} P(S_{n+\gamma-1}|\underline{J}_{n+\delta}) \tag{34}$$

and Equations (14) to (17). The value $m = \gamma - 1$ sets up the solution of Equation (34) based on the probabilities of states $S_{n+\gamma-1} = (I_n, \ldots, I_{n+\gamma-1})$, in which, I_n is located in the last position. Hence, it reduces the number of backward recursions while keeping the complexity due to the forward recursion unchanged. This motivates us that should we set up a solution based on the

[2]Note that the computational complexity of the forward equation includes the cost of normalization as well ($2M^{\gamma}$). This is a common approach in this work in which the forward probabilities of *trellis states* which are stored to be used in the next time instant are always normalized. However, in practise we perform the required normalization of Equation (28) after we summed the multiplied forward and backward terms according to Equation (27). This only costs $2M$ operations as opposed to the original $2M^{\gamma}$ according to Equation (28).

a posteriori probabilities of a sequence larger than a state, we can reduce the number of the backward recursions and its complexity even further. In the following section, we present such solutions and examine if it leads to an smaller *overall* complexity as compared to the solution of Equation (34).

B.2 Alternative solutions

In this section, we reconsider the problem of finding the a posteriori probability of a symbol I_n, given the observed sequence $\underline{J}_{n+\delta}$ for the cases where $\delta \geq \gamma$. Motivated by the results and discussions presented in section III-B.1.case 2, we seek possibly more efficient solutions to calculate the required a posteriori symbol probability using the probability of a sequence larger than one state, i.e.,

$$(I_n, \ldots, I_{n+\gamma+L-1}),\ 0 < L \leq \delta - \gamma + 1,$$

we have,

$$P(I_n|\underline{J}_{n+\delta}) = \sum_{I_{n+1}} \sum_{I_{n+2}} \cdots \sum_{I_{n+\gamma+L-1}} P(I_n, \ldots, I_{n+\gamma+L-1}|\underline{J}_{n+\delta}) \tag{35}$$

where $P(\underline{I}^n_{n+\gamma+L-1}|\underline{J}_{n+\delta})$ is given by the following forward backward equation,

$$\begin{aligned} &P(I_n, \ldots, I_{n+\gamma+L-1}|\underline{J}_{n+\delta}) = \\ &\quad C\,.\,P(I_n, \ldots, I_{n+\gamma+L-1}|\underline{J}_{n+\gamma+L-1})\,.P(\underline{J}^{n+\gamma+L}_{n+\delta}|S_{n+\gamma+L-1}) \end{aligned} \tag{36}$$

The first term in Equation (36) is the a posteriori probability of a sequence of $L+\gamma$ symbols or L states ($\gamma, L > 0$), given the entire history of the received information from the channel. To calculate such a probability, in section III-A.2, we presented a set of solutions based on different constructions of an extended trellis structure for the source coder output with $M^{\gamma+L'}$, $0 \leq L' < L$ states in each stage. The second term in Equation (36) is a backward recursive term which is given by Equation (16). The number of backward recursions is equal to $\delta - \gamma - L + 1$, which as expected reduces with the increase of L and is always smaller than that of the solution in section III-B.1.case 2. However, increase of L results in a more complex forward term.

Using the results of sections III-A.1 and III-A.2, the overall complexity of the solution in Equation (35) is given by,

$$CC = 2(L - L' + 1)M^{L+\gamma} + 3(\delta - \gamma - L + 1)M^{\gamma+1} + (M+2)M^{\gamma+L'} + 2M \tag{37}$$

which includes the number of computations required to calculate the forward and backward terms, updating the forward probabilities of super states according to Equation (22) and multi-

plications and normalization required in Equation (36). The memory requirement includes the fixed amount of ROM required to store $M^{\gamma+1}$ transition probabilities and the RAM required is $O(M^{\gamma+L'})$.

This set of solutions can be optimized over the choices of L (sequence length) and L', $0 \leq L' < L$ (the state set extension factor for the forward term). Interestingly, using the results of section III-A.2, it can be shown that $L = 1$ optimizes the solution in terms of the computational complexity disregarding of the values of γ and δ for all $M \geq 2$. It is noteworthy that since $L = 1$ requires $L' = 0$ hence, $L = 1$ minimizes the memory requirement as well. Therefore, the optimum solution (for $\delta \geq \gamma$) in terms of the complexity, provided by the family of solutions of Equation (35) is based on the original source trellis and is given by,

$$P(I_n|\underline{J}_{n+\delta}) = \sum_{I_{n+1}} \sum_{I_{n+2}} \cdots \sum_{I_{n+\gamma}} P(I_n, \ldots, I_{n+\gamma}|\underline{J}_{n+\delta}) \tag{38}$$

where from (36) we have,

$$\begin{aligned} &P(I_n, \ldots, I_{n+\gamma}|\underline{J}_{n+\delta}) = \\ &\quad C \,.\, P(I_n, \ldots, I_{n+\gamma}|\underline{J}_{n+\gamma}) \,.P(\underline{J}_{n+\delta}^{n+\gamma+1}|S_{n+\gamma}) \end{aligned} \tag{39}$$

in which, using Equation (20), the forward term is given by,

$$P(\underline{I}_{n+\gamma}^{n}|\underline{J}_{n+\gamma}) = C \,.\, P(J_{n+\gamma}|I_{n+\gamma}) \,.\, P(I_{n+\gamma}|S_{n+\gamma-1}) \,.\, P(S_{n+\gamma-1}|\underline{J}_{n+\gamma-1}) \tag{40}$$

and the backward term is calculated using Equations (16) and (17). The probabilities of states in the forward term are then updated by the following,

$$P(S_{n+\gamma}|\underline{J}_{n+\gamma}) = \sum_{I_n \in \mathcal{J}} P(\underline{I}_{n+\gamma}^{n}|\underline{J}_{n+\gamma}) \tag{41}$$

From Equation (37), the overall complexity of this solution is given by,

$$CC = (3\delta - 3\gamma + 5)M^{\gamma+1} + 2M^{\gamma} + 2M. \tag{42}$$

Now, it remains to compare the above solution (based on $L = 1, L' = 0$) with that presented in section III-B.1.case 2. Examining Equations (42) and (31), indicates that the current solution maintains a lower complexity. Although both solutions are based on the same original source trellis, their distinction stems from using different forward recursive (and updating) equations (Equations (40) and (41) vs. Equation (14)). This in turn leads to the reduction of the backward term by an additional step as seen comparing Equations (39) and (28) for $m = \gamma - 1$.

IV. MAP Decoding: Problem Statement and Solutions

A. Maximum A Posteriori Symbol Decoder

An instantaneous symbol MAP decoder exploiting the residual redundancies in the form of a first-order Markov model was presented in [33]. Later in [34], a decoder that accommodates a certain delay in the decoding process was proposed for the same problem setup. Here, we present an optimal symbol MAP decoder when the residual redundancies are captured with a γ-order Markov model and a delay of δ is allowed in the decoding process.

The symbol MAP decoder receives the sequence $\underline{J}_{n+\delta}$ and determines the most probable transmitted symbol. Next, it outputs the corresponding codeword. We have,

$$\begin{aligned} \hat{\mathbf{x}}_n &= E[\mathbf{X}_n | I_n = \hat{i}_n] \\ \hat{i}_n &= \arg \max_{I_n \in \mathcal{J}} P(I_n | \underline{J}_{n+\delta}) \end{aligned} \tag{43}$$

The required a posteriori probability of the symbol I_n in Equation (43) can be efficiently calculated as described in section III-B. The performance of this decoder is studied in section V where it is referred to as the MAP decoder.

The presented MAP decoder uses a codebook identical to that of the encoder. Alternatively, we can use the decoder codebook corresponding to the asymptotically optimum MMSE decoding algorithm with the MAP decoder. In this case a sequence is decoded such that,

$$\hat{\underline{i}}_{n+\delta}^{n-\tau} = \arg \max_{\underline{I}_{n+\delta}^{n-\tau} \in \mathcal{J}^{\tau+\delta+1}} P(\underline{I}_{n+\delta}^{n-\tau} | \underline{J}_{n+\delta}) \tag{44}$$

using Equations (18) or (20). Next, the source decoder reproduces,

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n | \underline{I}_{n+\delta}^{n-\tau} = \hat{\underline{i}}_{n+\delta}^{n-\tau}]$$

at the output. We refer to this technique as the AOMAP decoder and present its performance in section V.

B. Maximum A Posteriori Sequence Decoder

A sequence MAP decoder exploiting the residual redundancies in the form of a first-order Markov model was presented in [22] for source decoding over noisy channels. Later in [33], a similar but optimal decoder was proposed. Here we present an optimal sequence MAP decoder when the residual redundancies are captured with a γ-order Markov model.

The sequence MAP decoder receives the sequence $\underline{J}_T$ and determines the most probable transmitted sequence,

$$\hat{\underline{i}}_T = \arg \max_{\underline{I}_T \in \mathcal{J}^T} P(\underline{I}_T | \underline{J}_T) \tag{45}$$

Using the same trellis structure as described in the previous section and considering the memoryless property of the channel as well as the Markov model for the source redundancy, it is straightforward to see that the Equation (45) is equivalent to (see Appendix-C for a proof),

$$\hat{\underline{i}}_T = \arg \max_{\underline{I}_T \in \mathcal{J}^T} \left[\sum_{k=2}^{T} \log[P(J_k|I_k)P(I_k|S_{k-1})] + \log[P(J_1|I_1)P(S_1)] \right] \tag{46}$$

where $S_1 \triangleq (I_1, 0, \ldots, 0)$. The sequence MAP decoder in Equation (46) can be implemented using the well-known Viterbi algorithm. We use the same trellis structure as defined in section III-A.1 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 optimum sequence MAP decoder, according to Equation (46), requires to receive the whole sequence $\underline{J}_T$ to decode the corresponding sequence $\hat{\underline{i}}_T$ and hence, imposes a large delay. However, to limit the delay to a certain value, at each time instant, we identify the state with the maximum metric and decode the symbol at delay δ on the surviving path reaching that state accordingly. Subsequently, the correpsonding codeword is reproduced at the source decoder output. We refer to this decoder as the Sequence MAP (SMAP) decoder and will examin its performance in section V.

Given that the values $\log P(J|I)$ received from the channel are available, the computational complexity of the SMAP algorithm per time step is given by,

$$CC = 3M^{\gamma+1} + M^{\gamma} \tag{47}$$

which includes the computations required for updating the state metrics and selecting the one with the largest value.

The decoder codebook in the presented SMAP decoder is the same as the encoder codebook. Alternatively, we can use the decoder codebook corresponding to the asymptotically optimum MMSE decoding algorithm with the SMAP decoder. In this case a sequence is decoded which in turn outputs one of the decoder codewords $E[\mathbf{X}_n | \underline{I}_{n+\delta}^{n-\tau} = \hat{\underline{i}}_{n+\delta}^{n-\tau}]$. We refer to this technique as the AOSMAP decoder and present its performance in section V.

f_A	-1.3822	0.3399	0.1772	0.6760	-0.6396	0.0719	-0.1386	0.6024	-0.4561	0.1560
f_B	1.381	-0.599	0.367	-0.700	0.359					
f_C	-1.7493	2.4263	-2.5733	1.6585	-0.7426	0.1644	0.3019	-0.1157	0.0350	0.0018

TABLE I

FILTER COEFFICIENTS OF THE SYNTHESIZED SOURCES

V. NUMERICAL RESULTS

To analyze the performance of the proposed MMSE decoders, we use a synthesized source similar to [22]. In addition to the fifth-order Gauss-Markov source from [22], two other tenth-order Gauss-Markov sources are used whose coefficients have been picked from a speech LPC database. Each ten LPC coefficient set represents the short-time spectral information of speech within 20ms. The source samples are given by,

$$x_k = \sum_{i=1}^{\gamma_s} f(i).x_{k-i} + e_k \tag{48}$$

where E_k is a Gaussian i.i.d. random variable, γ_s is the order of the synthesizing filter and the corresponding coefficients $f(i)$ are given in Table I. The source sample vector $\mathbf{X}_n = [X_{(n-1)N+1}, \ldots, X_{nN}]$ is quantized with an M point N dimensional VQ producing the symbol I_n. At different redundancy model orders γ, the value $R(M, \gamma)$ in bits defined as,

$$R(M, \gamma) \triangleq \log_2 M - H(I_n|S_{n-1}) \tag{49}$$

where $S_n = [I_n, \ldots, I_{n-\gamma+1}]$, provides an indication of the redundancy to be exploited and hence, the gains to be achieved. Table II presents the amount of $R(M, \gamma)$ for the selected synthesized sources at different values of γ when the source is quantized by a 3-bit LBG scalar quantizer. As given in Table II, for source A, the redundancy due to the non-uniform distribution ($\gamma = 0$) is 0.25 bits. The redundancy exploited by means of a first, second and third order Markov model is 1.16, 1.48 and 1.67 bits respectively. In Table III, the normalized auto-correlation of the source samples at different delays are also presented. In the followings, we investigate the performance of the decoders presented in the previous sections. Six source decoders are considered: (i) asymptotically optimum MMSE (AOMMSE) decoder (ii) simplified MMSE (MMSE) decoder which uses identical encoder and decoder codebooks (iii) Maximum A Posteriori symbol decoder

Redundancy Order γ	0	1	2	3
R_A	0.254	1.162	1.482	1.667
R_B	0.319	1.182	1.327	1.388
R_C	0.439	0.540	0.867	1.248

TABLE II

REDUNDANCY OF THE SOURCE, $R(M,\gamma)$ (IN BITS), AT DIFFERENT REDUNDANCY MODEL ORDERS γ, $(M = 8, N = 1)$.

Delay δ	0	1	2	3
ρ_A	1.000	0.849	0.553	0.159
ρ_B	1.000	0.868	0.599	0.280
ρ_C	1.000	0.374	-0.417	-0.098

TABLE III

NORMALIZED AUTOCORRELATION OF THE SOURCE SAMPLES AT DIFFERENT DELAYS

(MAP) (iv) AOMAP decoder which selects the codeword with the Maximum A Posteriori probability from the codebook corresponding to the AOMMSE decoder (v) the Sequence MAP decoder (SMAP) (vi) and the AOSMAP decoder which is the Sequence MAP decoder which uses the decoder codebook of the AOMMSE decoder. We begin with the performance comparison of the instantaneous decoders ($\delta = 0$) over a Binary Symmetric Channel and we proceed to analyze the effect of delay, performance with a channel (decoder) with soft outputs, the effect of redundancy type and the effect of quantizer bit-rate.

A. Basic Comparison of the Decoders

In this section, we present a performance comparison of the instantaneous decoders ($\delta = 0$). Figure 2 demonstrates the performance of the instantaneous AOMMSE decoder for $\tau = \gamma$ (dotted lines) and the MMSE decoder (solid lines), for transmission of source A over a Binary Symmetric Channel when different levels of residual redundancy is exploited at the receiver ($\gamma = 1, 2, 3$). As mentioned before, for $\gamma = 0$ (and $\tau = 0$) both schemes collapse to the basic MMSE decoder of Equation (25). The performance of the basic MMSE decoder ($\gamma = 0$) with Equal symbol

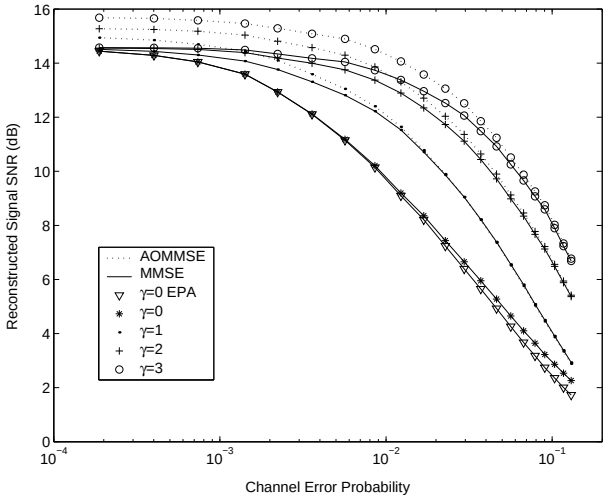

Fig. 2. Performance of the instantaneous ($\delta = 0$) AOMMSE (for $\tau = \gamma$) and MMSE decoders for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over a Binary Symmetric Channel when different levels of residual redundancy is exploited at the decoder. Note that for $\gamma = 0$ and $\gamma = 0$ EPA the curves of AOMMSE and MMSE decoding have overlapped.

Probability Assumption (EPA) in which case no a priori information is used in the decoding process ($P(I = i) = \frac{1}{M}, \forall i \in \mathcal{J}$) is provided as a baseline for comparison.

For the AOMMSE decoder, Figure 2 shows that using a redundancy model of order $\gamma = 2$ or $\gamma = 3$ provides a gain as high as 2.5dB or 4dB respectively compared to the case where the redundancy is modeled with a first-order Markov model. Using the simplified MMSE decoder, similar gains are achievable, however at lower bit error rates, the performance is upperbounded by that provided at the encoder output. As mentioned before, in such cases the AOMMSE decoding provides a finer reconstruction of the source samples. This is due to using a larger decoder codebook which exploits the dependencies between the source coder output symbols. The performance of the corresponding AOMAP and MAP decoders are presented in Figure 3. In Figure 4, the performance of a selected set of instantaneous AOMMSE, MMSE, AOMAP and MAP decoders are redrawn for comparison. It is observed that the MMSE decoders constantly outperform the MAP decoders with gains as high as 1.4dB.

Figure 5 compares the performance of the MAP symbol decoder with that of the Sequence MAP decoder for transmission of source A over a Binary Symmetric Channel. It is observed that the SMAP algorithm, although suboptimal in the sense of minimizing the symbol probability of error, performs very closely to the MAP algorithm in the MSE sense. For any given delay of

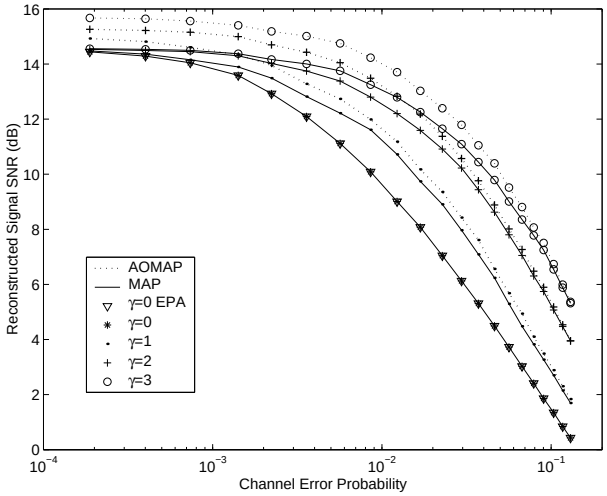

Fig. 3. Performance of the instantaneous ($\delta = 0$) AOMAP (for $\tau = \gamma$) and MAP decoders for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over a Binary Symmetric Channel when different levels of residual redundancy is exploited at the decoder. Note that for $\gamma = 0$ and $\gamma = 0\,\mathrm{EPA}$ the curves of AOMAP and MAP decoding have overlapped.

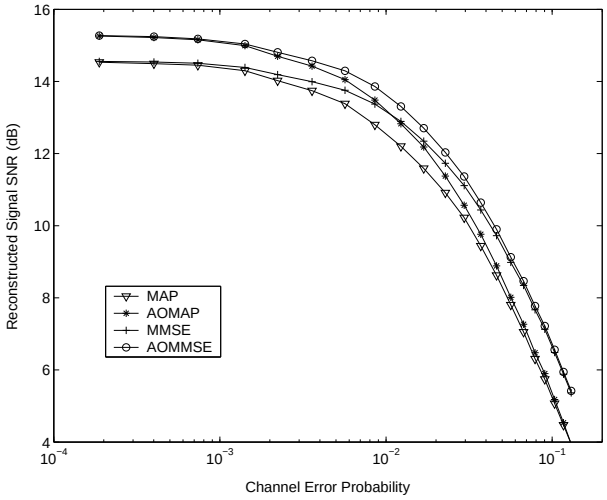

Fig. 4. Performance of the instantaneous ($\delta = 0$) MAP, AOMAP, MMSE and AOMMSE decoders for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over a Binary Symmetric Channel when the redundancy order $\gamma = 2$ and $\tau = \gamma$.

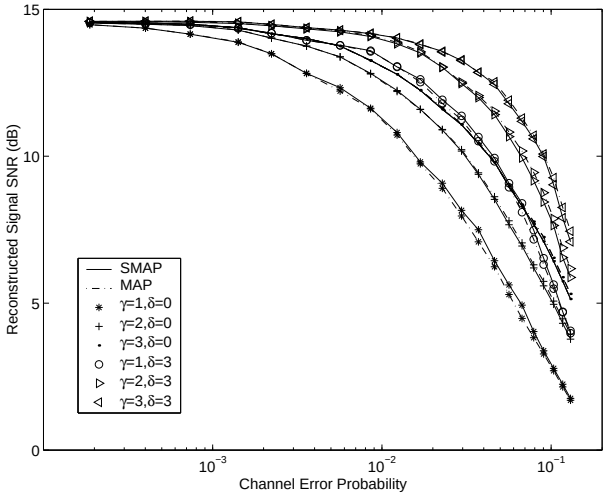

Fig. 5. Performance of the SMAP and MAP decoders for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over a Binary Symmetric Channel when different levels of residual redundancy is exploited at the decoder, $\delta = 1, 3$.

δ and redundancy order γ, similar observations are made in other cases, when comparing the AOMAP and the AOSMAP decoders (with the same τ) or the MAP and the SMAP decoders. Consequently, we will only discuss the performance of the MAP and AOMAP algorithms in the following sections. As seen in section IV, the SMAP and the MAP algorithm can be implemented with a comparable complexity. A more precise complexity comparison depends on the design parameters such as δ and γ and the actual decoder implementation. The AOSMAP decoder maintains a lower level of complexity as compared to the AOMAP decoder.

B. Effect of Delay

To demonstrate the effect of delay, Figure 6 depicts the performance of the MMSE decoder for reconstruction of the source A over a Binary Symmetric Channel at different delays ($\delta = 0, 1, 2, 3$) for the two scenarios of redundancy order $\gamma = 1$ and $\gamma = 3$. The curve corresponding to the basic MMSE decoder ($\gamma = 0$) of Equation (25) provides a baseline for comparison. Also the performance of the MAP decoder in similar scenarios are provided in Figure 7. It is observed that for the case of transmission of source A over a noisy channel, a delay of $\delta = 3$ allows the decoder to capture almost all of the redundancy in the future samples. The gains achieved in this case are higher than 3.5dB, 3dB and 2.5dB for $\gamma = 1$, $\gamma = 2$ and $\gamma = 3$ respectively when compared with the corresponding instantaneous decoding schemes.

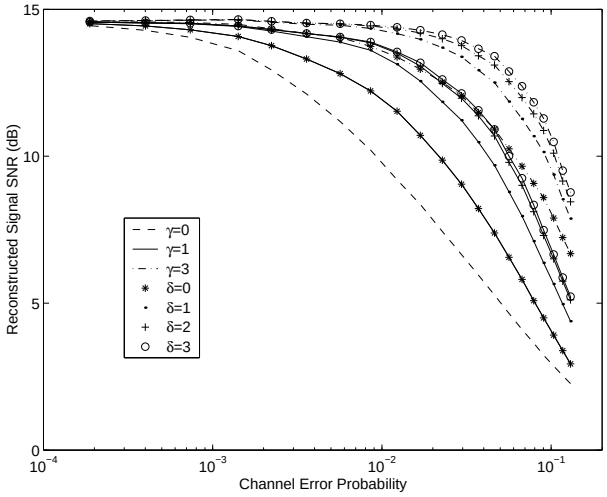

Fig. 6. Performance of the MMSE decoder for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over a Binary Symmetric Channel when different delays are allowed ($\delta = 0, 1, 2, 3$) and the residual redundancy is exploited with a $\gamma = 1, 3$ order Markov model.

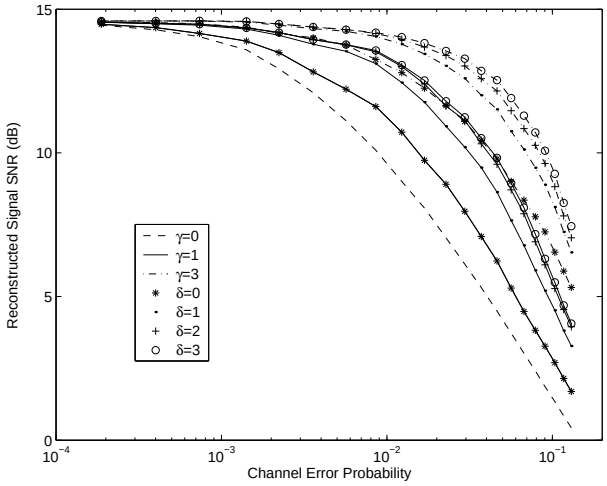

Fig. 7. Performance of the MAP decoder for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over a Binary Symmetric Channel when different delays are allowed ($\delta = 0, 1, 2, 3$) and the residual redundancy is exploited with a $\gamma = 1, 3$ order Markov model.

C. Performance Using A Soft Output Channel (Decoder)

Recently, channel decoding techniques using the soft channel information has found increasing attention in different applications for their improved performance. In techniques such as turbo decoding, iterative decoding or soft output Viterbi algorithm, soft outputs are readily available at the output of the channel decoder as well. The decoders proposed in this work are able

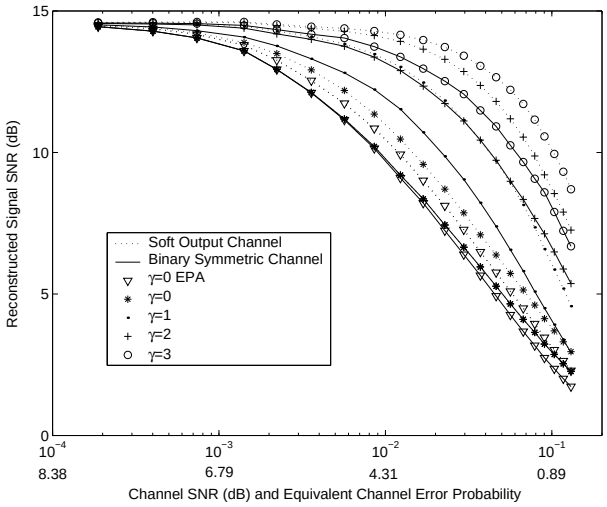

Fig. 8. Performance of the MMSE decoder for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over the Soft Output Channel and the Binary Symmetric Channel when different levels of source redundancy is exploited at the decoder, $\delta = 0$.

to exploit the soft output information and the source a priori information for effective source decoding. Alternatively, with appropriate considerations the proposed MAP (SMAP) decoders can be used for effective channel decoding using the soft channel information and assisted with the source a priori information.

To indicate the possible performance improvement due to using the soft output of the channel (decoder), Figure 8 compares the performance of the instantaneous MMSE decoder for reconstruction of source A transmitted over the Soft Output Channel and the Binary Symmetric Channel. Figure 9 provides the same comparison when a delay of $\delta = 3$ is allowed in the decoding process. Alternatively, Figures 10 and 11 depict the same performance results when MAP decoding is used. It is observed that if channel (decoder) soft outputs are available gains as high as $2-3$dB can be achieved. As the decoding schemes become stronger i.e. δ and γ are increased the maximum gains achieved move more towards the low channel SNRs or higher probabilities of error.

D. Effect of Redundancy

In this section, we study the effect of the type of source redundancy in the achievable gains using the proposed techniques. As well, we examine the effectiveness of the measures of redundancy as discussed before. We consider the instantaneous MMSE reconstruction of the sources A, B and C

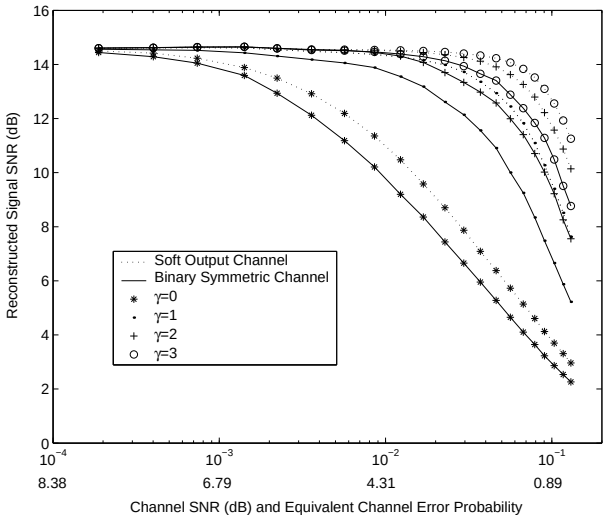

Fig. 9. Performance of the MMSE decoder for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over the Soft Output Channel and the Binary Symmetric Channel when different levels of source redundancy is exploited at the decoder, $\delta = 3$.

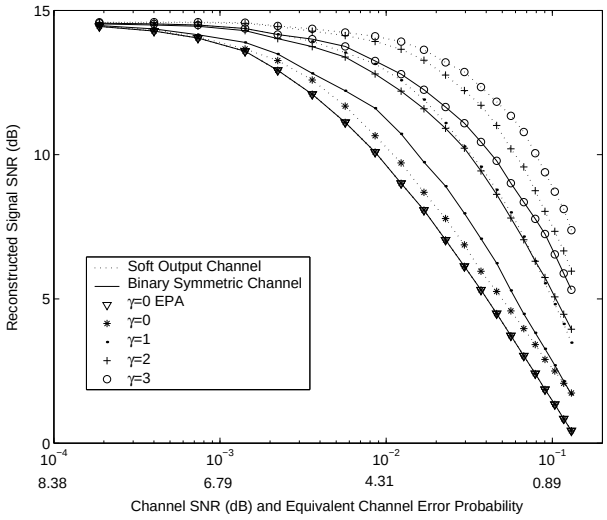

Fig. 10. Performance of the MAP decoder for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over the Soft Output Channel and the Binary Symmetric Channel when different levels of source redundancy is exploited at the decoder, $\delta = 0$. Note that the curves corresponding to BSC with $\gamma = 0$, $\gamma = 0\,\mathrm{EPA}$ and SOC with $\gamma = 0\,\mathrm{EPA}$ have overlapped.

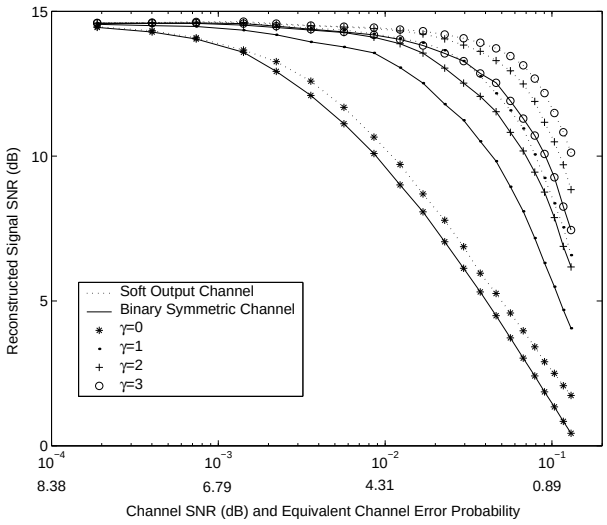

Fig. 11. Performance of the MAP decoder for transmission of the Gauss-Markov source A ($M = 8$, $N = 1$) over the Soft Output Channel and the Binary Symmetric Channel when different levels of source redundancy is exploited at the decoder, $\delta = 3$.

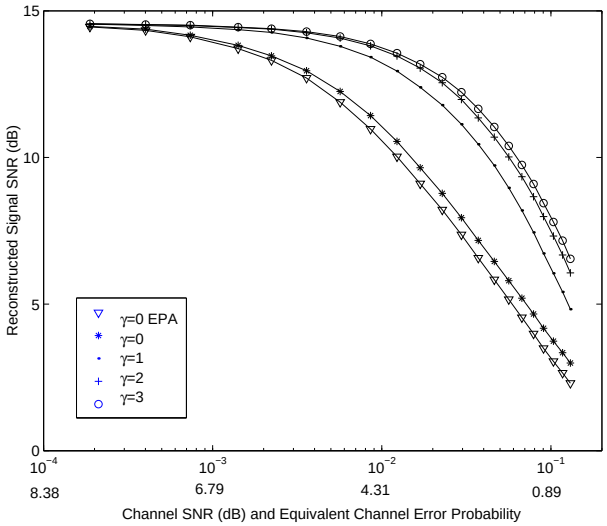

Fig. 12. Performance of the instantaneous MMSE decoder for transmission of the Gauss-Markov source B ($M = 8$, $N = 1$) over the Soft Output Channel when different levels of source redundancy is exploited.

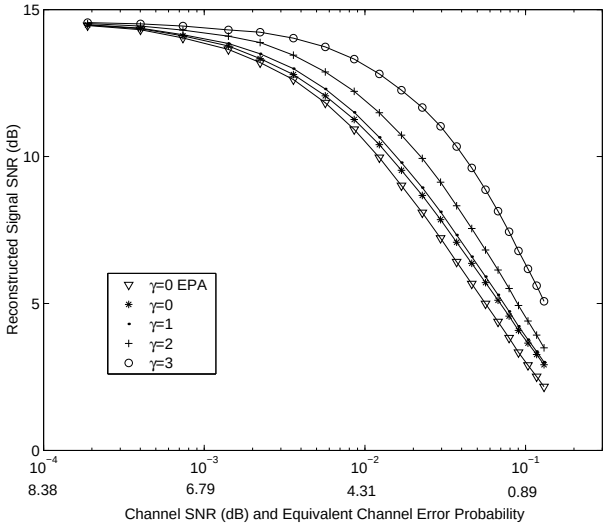

Fig. 13. Performance of the instantaneous MMSE decoder for transmission of the Gauss-Markov source C ($M = 8$, $N = 1$) over the Soft Output Channel when different levels of source redundancy is exploited.

over the Soft Output Channel as given in Figures 8,12 and 13 respectively. From these figures, it is observed that the amount of redundancy $R(M, \gamma)$, as defined in Equation (49) and provided in Table II correlates well with the achieved gains. On the other hand, the source auto-correlation as given in Table III does not seem to be a suitable indicator of the possible gains. This is in line with the observations in [22].

E. Effect of Quantizer Bitrate

Figure 14 depicts the performance of the instantaneous MMSE decoder for transmission of source A quantized with an $M = 4, 8, 16$ point quantizer over the Soft Output Channel model. Since the higher rate quantizers are more sensitive to the channel errors, therefore the effectiveness of the proposed decoder is more significant in such cases. Specifically, the gains at low error rates are noticeable.

VI. Conclusions

A family of solutions for the asymptotically optimum MMSE reconstruction of a source over a memoryless noisy channel is presented when the redundancy in the source encoder output stream is exploited in the form of a γ-order Markov model ($\gamma \geq 1$) and a delay of δ, $\delta > 0$, is allowed in the decoding process. Considering the same problem setup, we also present a simplified MMSE

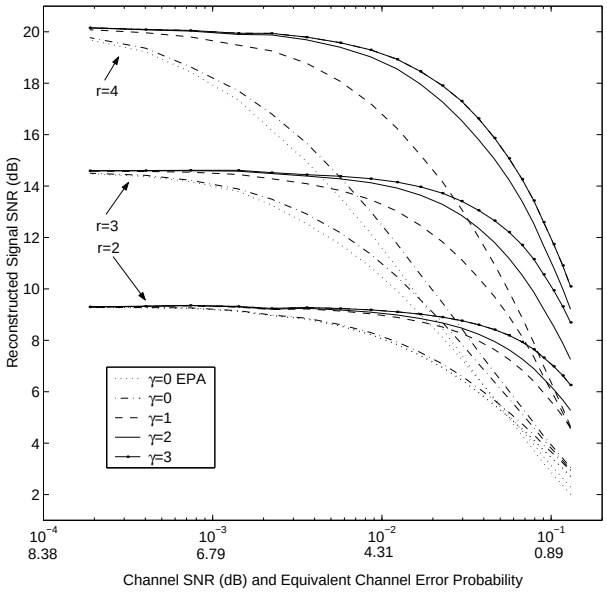

Fig. 14. Performance of the instantaneous MMSE decoder for transmission of the Gauss-Markov source A (quantized with rates $r = 2, 3$ and 4 bits, $N = 1$) over the Soft Output Channel when different levels of source redundancy is exploited, $M = 2^r$, $N = 1$.

decoder as well as several other Maximum A Posteriori symbol and sequence decoders. In each case, we investigate the alternative solutions and optimize them for the smallest computational complexity.

The numerical results and analysis demonstrate the effectiveness of the stronger models (higher Markov order, γ) to capture the residual redundancy. The MMSE-based decoders outperform their equivalent MAP-based decoders. As expected, the asymptotically optimum MMSE (AOMMSE) decoder provides the best performance among the presented decoders. The simplified MMSE decoder has a smaller decoder codebook and a lower complexity, which is comparable to that of the SMAP decoder. The sequence MAP decoder and the symbol MAP decoder maintain the same level of performance. The AOSMAP decoder provides a lower complexity alternative to the AOMMSE decoder at the price of a certain loss in performance.

The possible future research in this direction includes the design of Channel Optimized Vector Quantizers based on the proposed decoders and more efficient approximate algorithms to the presented MMSE decoders.

Appendix

A. Proofs for section III-A.1

The a posteriori probabilities of states within the source trellis diagram, described in section III-A.1, are calculated by the following forward backward equation as detailed below,

$$\begin{aligned} P(S_n|\underline{J}_{n+\delta}) &= C_1.P(S_n, \underline{J}_n, \underline{J}_{n+\delta}^{n+1}) \\ &= C.P(S_n|\underline{J}_n).P(\underline{J}_{n+\delta}^{n+1}|S_n, \underline{J}_n) \\ &= C.P(S_n|\underline{J}_n).P(\underline{J}_{n+\delta}^{n+1}|S_n). \end{aligned} \tag{50}$$

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,

$$\begin{aligned} P(S_n|\underline{J}_n) &= C_1.P(S_n, \underline{J}_n) \\ &= C_1.P(\underline{J}_{n-1}).P(S_n|\underline{J}_{n-1}).P(J_n|S_n, \underline{J}_{n-1}) \\ &= C.P(J_n|I_n).P(S_n|\underline{J}_{n-1}) \\ &= C.P(J_n|I_n).\sum_{S_{n-1}} P(S_n|S_{n-1}, \underline{J}_{n-1}).P(S_{n-1}|\underline{J}_{n-1}) \\ &= C.P(J_n|I_n).\sum_{S_{n-1}} P(S_n|S_{n-1}).P(S_{n-1}|\underline{J}_{n-1}), \end{aligned} \tag{51}$$

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 follows,

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

Using Equations (1) and (3) 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,

$$\begin{aligned} P(\underline{J}_{n+\delta}^{n+1}|S_n) &= \sum_{I_{n+1}} P(J_{n+1}|I_{n+1})P(I_{n+1}|S_n) \sum_{I_{n+2}} P(J_{n+2}|I_{n+2})P(I_{n+2}|S_{n+1}) \dots \\ &\quad \sum_{I_{n+\delta}} P(J_{n+\delta}|I_{n+\delta})P(I_{n+\delta}|S_{n+\delta-1}) \\ &= \sum_{I_{n+1}} P(J_{n+1}|I_{n+1})P(I_{n+1}|S_n)P(\underline{J}_{n+\delta}^{n+2}|S_{n+1}) \end{aligned} \tag{52}$$

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.

B. Proofs for section III-A.2

The Equation (20) provides the probability of a sequence of symbols within the structure of the trellis, given the entire history of the received signals. The derivation of this Equation is presented below. Note that, $L > L' \geq 0$ and for $L' = 0$, this collapses to the case with the original source trellis presented in Equation(18).

$$\begin{aligned} P(\underline{I}_n^{n-(\gamma+L)+1}|\underline{J}_n) &= C_1.P(\underline{I}_n^{n-(\gamma+L)+1}, \underline{J}_n) \\ &= C_1.P(SS_{n-(L-L')}, \underline{I}_n^{n-(L-L')+1}, \underline{J}_{n-(L-L')}, \underline{J}_n^{n-(L-L')+1}) \\ &= C.P(SS_{n-(L-L')}|\underline{J}_{n-(L-L')}).P(\underline{I}_n^{n-(L-L')+1}|SS_{n-(L-L')}, \underline{J}_{n-(L-L')}) \\ &\quad .P(\underline{J}_n^{n-(L-L')+1}|SS_{n-(L-L')}, \underline{J}_{n-(L-L')}, \underline{I}_n^{n-(L-L')+1}) \end{aligned}$$

Using the memoryless property of the channel and the Markovian property of the source, this is simplified to,

$$\begin{aligned} &P(\underline{I}_n^{n-(\gamma+L)+1}|\underline{J}_n) = C\,.\,P(\underline{J}_n^{n-(L-L')+1}|\underline{I}_n^{n-(L-L')+1}). \\ &\quad P(\underline{I}_n^{n-(L-L')+1}|SS_{n-(L-L')})\,.\,P(SS_{n-(L-L')}|\underline{J}_{n-(L-L')}) = \\ &\quad C\,.\left[\prod_{k=0}^{L-L'-1} P(J_{n-k}|I_{n-k})P(I_{n-k}|SS_{n-k-1})\right].P(SS_{n-(L-L')}|\underline{J}_{n-(L-L')}) \end{aligned} \tag{53}$$

where in Equation (53), $C = P(\underline{J}_{n-(L-L')})/P(\underline{J}_n)$. The overall computational complexity of the Equation (53) is given by,

$$CC = 2(L - L' + 1)M^{\gamma+L} + (M + 2)M^{\gamma+L'} \tag{54}$$

Now, we show that for a fixed value of $L, L > 0$, increasing the value of L', $0 \leq L' < L$, or the state set extension factor, reduces this complexity. The case of $L = 1$ requires $L' = 0$ and is trivial. For a given L, $L > 1$, $\forall L'$, $0 < L' < L$, we have,

$$\begin{aligned} CC(L, L') - CC(L, L' - 1) &= -2M^{\gamma+L} + M^{\gamma+L'-1}(M^2 + M - 2) \\ &= M^{\gamma+L'-1}(-2M^{L-L'+1} + M^2 + M - 2) \\ &\leq M^{\gamma+L'-1}(-M^2 + M - 2) < 0 \end{aligned}$$

C. Proof for section IV-B

The Sequence MAP decoder is derived as follows,

$$\hat{\underline{i}}_T = \arg \max_{\underline{I}_T \in \mathcal{J}^T} P(\underline{I}_T | \underline{J}_T) \tag{55}$$

in which,

$$\begin{aligned} P(\underline{I}_T|\underline{J}_T) &= C.P(\underline{J}_T|\underline{I}_T)P(\underline{I}_T) = C.P(\underline{J}_T|\underline{I}_T)P(\underline{S}_T) \\ &= C.\left[\prod_{k=1}^{T} P(J_k|I_k)\right] . \left[\prod_{k=2}^{T} P(S_k|S_{k-1})P(S_1)\right] \\ &= C.\prod_{k=2}^{T} P(J_k|I_k)P(I_k|S_{k-1})\, P(J_1|I_1)P(S_1), \end{aligned} \tag{56}$$

and $C = \frac{1}{P(\underline{J}_T)}$. Considering that the logarithm function is monotonically increasing, we have,

$$\begin{aligned} \hat{\underline{i}}_T &= \arg \max_{\underline{I}_T \in \mathcal{J}^T} P(\underline{I}_T|\underline{J}_T) = \arg \max_{\underline{I}_T \in \mathcal{J}^T} \log P(\underline{I}_T|\underline{J}_T) \\ &= \arg \max_{\underline{I}_T \in \mathcal{J}^T} \left[\sum_{k=2}^{T} \log[P(J_k|I_k)P(I_k|S_{k-1})] + \log[P(J_1|I_1)P(S_1)]\right] \end{aligned} \tag{57}$$

where $S_1 \triangleq (I_1, 0, \ldots, 0)$.

References

[1] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.,* vol. 27, pp. 379-423 and 623-656, 1948.

[2] J. W. Modestino, D. G. Daut, and A. L. Vickers, "Combined source channel coding of images using the block cosine transform," *IEEE Trans. Commun.,* vol. COM-29, pp. 1262-1274, Sept. 1981.

[3] C. C. Moore and J. D. Gibson, "Self-orthogonal convolutional coding for the DPCM-AQB speech encoder," *IEEE Trans. Commun.,* vol. COM-32, Aug. 1984.

[4] 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, No. 6, pp. 880-890, June 2000.

[5] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory,* vol. 43, No. 5, pp. 1412-1424, Sept. 1997.

[6] A. Méhes and K. Zeger, "Performance of quantizers on noisy channels using structured families of codes," *IEEE Trans. Inform. Theory,* vol. 46, No.7, pp. 2468 -2476, Nov. 2000.

[7] "TDMA radio interface, Enhanced full-Rate speech codec," TIA/EIA PN-3467, Feb. 1996.

[8] C.-E. Sundberg, "The effect of single bit errors in standard nonlinear PCM systems," *IEEE Trans. Commun.,* vol. COM-24, pp. 1062-1064, June 1976.

[9] S. Gadkari and K. Rose, "Unequally protected multistage vector quantization for time-varying CDMA channels," *IEEE Trans. Commun.,* vol. 49, No. 6, pp. 1045-1054, June 2001.

[10] I. Kozintsev and K. Ramchandran, "Robust image transmission over energy-constrained time-varying channels using multiresolution joint source-channel coding," *IEEE Trans. Signal Processing,* vol. 46, No. 4, pp. 1012-1026, April 1998.

[11] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory,* vol. 44, No. 6, Oct. 1998.

[12] A. J. Kurtenbach and P. A. Wintz, "Quantizing for noisy channels," *IEEE Trans. Commun.,* vol. COM-17, pp. 291-302, Apr. 1969.

[13] 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.

[14] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electron. Eng. Japan,* vol. 67-B, No. 4, pp. 39-47, 1984.

[15] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. Inform. Theory,* vol. 37, No.1, pp. 155-160, Jan. 1991.

[16] J. G. Dunham and R. M. Gray, "Joint source and channel trellis encoding," *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 516-519, July 1981.

[17] 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.

[18] 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.

[19] Z. Guang-Chong and F. I. Alajaji, "Soft-decision COVQ for turbo-coded AWGN and Rayleigh fading channels," *IEEE Communications Letters,* vol. 5, No. 6, pp. 257-259, June 2001.

[20] 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 Science, pp. 347-396, 1995.

[21] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Vector Quantization for Speech Transmission," in *Speech Coding and Synthesis,* W. Kleijn and K. Paliwal, Eds. New York: Elsevier Science, pp. 311-345, 1995.

[22] K. Sayood, J. C. Brokenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Trans. Commun.,* vol.39, No.6, pp. 838-845, 1991.

[23] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.,* vol. 43, No. 9, Sept. 1995.

[24] K. Sayood, L. Fuling and J. D. Gibson, "A constrained joint source/channel coder design," *IEEE Trans. Selec. Areas Commun.,* vol. 12, No. 9, pp. 1584-1593, Dec. 1994.

[25] F. I. Alajaji, N. Phamdo and T. E. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech and Audio Proc.,* vol. 4, No. 5, Sept. 1996.

[26] A. Ruscitto and E. M. Biglieri, "Joint source and channel coding using turbo codes over rings," *IEEE Trans. Commun.,* vol. 46, No. 8, pp. 981-984, Aug. 1998.

[27] T. Fazel and T. E. Fuja, "Joint source-channel decoding of block-encoded compressed speech," in *Proc. Confs. Information Sciences and Systems,* pp. FA5-1-FA5-6, Mar. 2000.

[28] T. Fingscheidt and P. Vary, "Softbit speech decoding: A new approach to error concealment," *IEEE Trans. on Speech and Audio Proc.,* vol. 9, No. 3, Mar. 2001.

[29] F. Lahouti and A. K. Khandani, "Approximating and exploiting the residual redundancies- Applications to efficient reconstruction of speech over noisy channels," *Proc. IEEE Int. Confs. Acoust., Speech and Sig. Proc.,* vol.2, pp.721-724, Salt Lake City, UT, May 2001.

[30] M. Adrat, J. Spittka, S. Heinen and P. Vary, "Error concealment by near optimum MMSE-estimation of source codec parameters," *IEEE Workshop on Speech Coding,* pp. 84-86, 2000.

[31] F. Lahouti and A. K. Khandani, "An efficient MMSE source decoder for noisy channels," *Proc. Int. Symp. Telecommun.,* pp. 784-787, Tehran, Iran, Sept. 2001.

[32] F. Lahouti and A. K. Khandani, "Sequence MMSE source decoding over noisy channels using the residual redundancies", *Annual Allerton Conference on Communication, Control and Computing,* Illinois, Urbana-Champaign, October 2001.

[33] 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, 1994.

[34] 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, No.2, pp. 222-231, 1998.

[35] N. Phamdo, F. Alajaji and N. Farvardin, "Quantization of memoryless and Gauss-Markov sources over binary Markov channels," *IEEE Trans. Commun.,* vol. 45, No. 6, pp. 668-675, June 1997.

[36] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Trans. Inform. Theory,* vol. 45, No. 4, pp. 1293-1307, 1999.

[37] T. Hindelang, T. Fingscheidt, N. Seshadri and R. V. Cox, Combined source/channel (de-)coding: can a priori information be used twice? *Proc. IEEE Int. Confs. Coomun.,* vol. 3, pp. 1208-1212, 2000.

[38] T. Hindelang, T. Fingscheidt, N. Seshadri and R. V. Cox, "Combined source/channel (de-)coding: can a priori information be used twice?," *Proc. IEEE Int. Symp. Inform. Theory,* pp. 266, 2000.

[39] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE J. Select. Areas Commun.,* vol. 19, No. 9, pp. 1662-1670, Sept. 2001.

[40] M. Adrat, P.Vary and J. Spittka, "Iterative source-channel decoder using extrinsic information from softbit-source decoding," *Proc. IEEE Int. Confs. Acoust., Speech and Sig. Proc.,* vol. 4, pp. 2653-2656, Salt Lake City, UT, May 2001.

[41] T. Fingscheidt, T. Hindelang, R. V. Cox, and N. Seshadri, "Joint source-channel (de-)coding for mobile communications" to appear in *IEEE Trans. Commun.,* 2002.

[42] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Tran. on Inform. Theory,* vol. 20, pp. 284-287, Mar. 1974.

[43] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.,* vol. COM-28, pp. 84-95, 1980.