

Causal transmission of colored source frames over a packet erasure channel

Ying-zong Huang, Yuval Kochman & Gregory W. Wornell

Dept. of Electrical Engineering and Computer Science

Massachusetts Institute of Technology

Cambridge, MA 02139

{zong,yuvalko,gww}@mit.edu

Abstract

We propose a linear predictive quantization system for causally transmitting parallel sources with temporal memory (colored frames) over an erasure channel. By optimizing within this structure, we derive an achievability result in the high-rate limit and compare it to an upper bound on performance. The proposed system subsumes the well-known PCM and DPCM systems as special cases. While typically DPCM performs well without erasures and PCM suffers less with many erasures, we show that the proposed solution improves performance over both under all severities of erasures, with unbounded improvement in some cases.

1 Introduction

The coding and transmission of colored sources under a variety of constraints is a problem of practical interest. The linear predictive structure as exemplified by DPCM has long been favored in solutions to this problem as an intuitive and low-complexity way to exploit source memory.

Classically, scalar DPCM was used for audio and scanline image and video coding, due to its lower bit rate relative to, e.g., PCM while still allowing sequential processing. In recent years, inspired by the structure of interframe motion prediction in video coders, there has been renewed interest in the vector DPCM structure for which optimality as measured against fundamental bounds was proved in several settings. In [1], it is shown that for stationary Gaussian sources, the Gaussian rate-distortion bound is achievable at all rates by vector-quantized DPCM if non-causal pre- and post-filtering is allowed. Furthermore, an analysis contained in [2], if carried to the stationary setting, shows that for Gauss-Markov sources, DPCM with causal MMSE estimation is optimal among all causal systems.

In this paper, we take a line of inquiry similar to [2], but also consider channel erasures. In applications such as real-time internet video conferencing, it is more realistic to assume

This work was supported in part by Hewlett-Packard under the MIT/HP Alliance, and by a grant from Microsoft Research.

that both a causality constraint on coding and errors during transmission are present. The combination is particularly interesting because causality severely limits what can be done about the errors, e.g., channel coding is impossible. Although in this paper we do not find what the generally optimal coding system is for this set of constraints, we do determine what is optimal among linear predictive structures under a few assumptions. The linear predictive structure is chosen as a productive starting template for several reasons: (1) it was recognized very early that adverse effects of channel errors were reducible by altering predictors in DPCM [3, 4]; and (2) engineering practice in, e.g., video coding shows that good prediction (interframe coding) can be traded for robustness against errors (intraframe coding) within an overall predictive coding framework. However, let us note that while the work is motivated by practical applications, the solution is a mathematical demonstration of concept, albeit one that can still be a suggestive guide for designers.

After we formulate the problem more precisely in Section 2, we consider as a solution the linear predictive quantization system that generalizes both DPCM and PCM in Section 3, where we also show its behavior on simple sources. We optimize performance for general stationary Gaussian sources in the high-rate limit in Section 4, and give closed-form solutions for performance comparisons.

2 Problem formulation

The practical scenario we model is the following. Frames of data indexed by time are given to the encoder to be immediately encoded (e.g., quantized) at some fixed rate and then transmitted as packets across a link that erases packets at certain times. A decoder receives either the correct packet or knows the packet is erased, and immediately decodes and renders a reproduction frame, possibly aided by intelligent concealment making up for any missing data. We look for the encoder and decoder pair that optimizes the source-to-reproduction MSE distortion in the system. Our model for the problem follows.

2.1 Source model

Let $\{\mathbf{s}[t]\}_t = \{(s_1[t], \dots, s_N[t])\}_t$ be a vector of N parallel source sequences, which can be viewed as a sequence of frames. In our notation, motivated by the video coding application, the bracketed index t is the temporal dimension and the N vector components make up the (typically very large) spatial dimension. We assume the source is:

1. Spatially i.i.d.: For all t , $\{s_i[t]\}_t$ is independent of $\{s_j[t]\}_t$ whenever $i \neq j$, and they have the same distribution. We omit the spatial index i when referring to a representative scalar sequence $\{s_i[t]\}_t$ should there be no ambiguity.
2. Temporally autoregressive: Each scalar stream $\{s[t]\}_t$ is stationary AR(P), characterized by $s[t] = z[t] + \sum_{p=1}^P \alpha_p s[t-p]$, with white innovation process $z[t] \sim \mathcal{N}(0, \Sigma_z)$, and $\alpha_1, \dots, \alpha_P$ such that all roots of $1 - \sum_{p=1}^P \alpha_p z^{-p}$ satisfy $|z| < 1$.

Denote by $\Phi_s(f) = \Sigma_z / |1 - A(f)|^2$ the power spectral density of $s[t]$, where $A(f) = \sum_{p=1}^P \alpha_p e^{-j2\pi fp}$. The source variance is $\Sigma_s = \int_{-\frac{1}{2}}^{\frac{1}{2}} \Phi_s(f) df$. The distortion-rate function

$D_s(R)$ of the scalar process $\{s[t]\}_t$ at sufficiently high rates ($R \geq \frac{1}{2} \log \sup_f |1 - A(f)|^2$, where the Shannon Lower Bound is tight) is,

$$D_s(R) = \Sigma_z 2^{-2R}, \quad (1)$$

which equals the distortion-rate of the innovation process $\{z[t]\}_t$ (cf. [5], p. 233).

In the rest of this paper we always normalize $\Sigma_z = 1$, so that the P parameters $\alpha_1, \dots, \alpha_P$ (equivalently, $A(f)$) entirely characterize the source.

2.2 Channel model

At each time t , the channel $\mathcal{C}(\cdot)$ is capable of transmitting a packet of $\lceil NR \rceil$ bits. The channel has two states. The channel state is an i.i.d. process described by an erasure probability ϵ . In the no-erasure state occurring with probability $1 - \epsilon$, the input is reproduced exactly at the output. In the erasure state occurring with probability ϵ , the output is a special lost packet symbol \emptyset immediately recognizable by the decoder, but there is no feedback to the encoder.

2.3 Encoder and decoder models

At time t , the causal encoder takes current and past source vectors $\mathbf{s}[t], \mathbf{s}[t-1], \dots$ as input and produces a packet $\mathcal{E}_t = \text{ENC}(\mathbf{s}[t], \mathbf{s}[t-1], \dots)$ of $\lceil NR \rceil$ bits. The causal decoder takes current and past receptions $\mathcal{C}(\mathcal{E}_t), \mathcal{C}(\mathcal{E}_{t-1}), \dots$ and produces $\hat{\mathbf{s}}[t] = \text{DEC}(\mathcal{C}(\mathcal{E}_t), \mathcal{C}(\mathcal{E}_{t-1}), \dots)$, a reproduction of the newest source vector $\mathbf{s}[t]$.

2.4 Performance and objective

The average distortion per scalar sample between source $\mathbf{s}[t]$ and its reproduction $\hat{\mathbf{s}}[t]$ at time t is the MSE distortion averaged over the spatial dimension: $d(\mathbf{s}[t], \hat{\mathbf{s}}[t]) = \frac{1}{N} \sum_{i=1}^N (s_i[t] - \hat{s}_i[t])^2$. Define the time-averaged expected distortion to be:

$$D(R; \epsilon) \triangleq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E} \{d(\mathbf{s}[t], \hat{\mathbf{s}}[t])\},$$

where expectation is taken over all source and channel realizations. The objective of the problem is to minimize the *excess distortion ratio*

$$L(R; \epsilon) \triangleq D(R; \epsilon) / D_s(R),$$

for the supplied R and ϵ and $D_s(R)$ as in (1), by designing the encoder and decoder. We see later that when looking at the high-rate limit, it is more insightful to let R grow and ϵ shrink such that $\lambda \triangleq \epsilon / 2^{-2R}$ is a fixed constant. In that case, the figure of merit becomes

$$L^\infty(\lambda) \triangleq \lim_{R \rightarrow \infty} L(R; \lambda 2^{-2R}) = \lim_{R \rightarrow \infty} D(R; \lambda 2^{-2R}) / 2^{-2R}, \quad (0 < \lambda < \infty). \quad (2)$$

3 Preliminary insights

As discussed in Section 1, we consider the problem within the class of linear predictive quantization systems for reasons of tractability and the importance that such systems play in engineering practice.

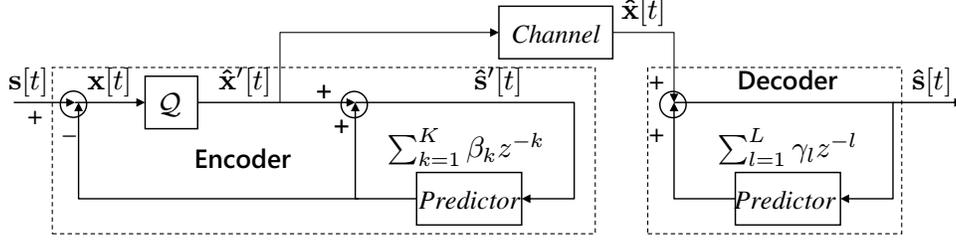


Figure 1: A causal, linear predictive quantization system model. \mathcal{Q} is a quantizer coding the N -vector $\mathbf{x}[t]$.

3.1 Linear predictive quantization system

While more general structures are possible, Fig. 1 shows the linear predictive encoder and decoder structure we consider in this paper.

Since erasures occur at the packet level, we suppress details relating to bitstream encoding and decoding for the channel. With that abridgement, $\hat{\mathbf{x}}'[t]$ and $\hat{\mathbf{x}}[t]$ are respectively the channel input and output of the system, with $\hat{\mathbf{x}}[t] = \hat{\mathbf{x}}'[t]$ when there is no erasure and $\hat{\mathbf{x}}[t] = \emptyset$ otherwise.

The quantizer \mathcal{Q} takes the differential signal $\mathbf{x}[t] = \mathbf{s}[t] - \sum_{k=1}^K \beta_k \hat{\mathbf{s}}'[t - k]$ as input and produces as output $\hat{\mathbf{x}}'[t] \in \hat{\mathcal{X}}_t \subseteq \mathbf{R}^N$ at rate R per scalar sample ($H(\hat{\mathbf{x}}'[t])/N = R$). We model \mathcal{Q} as a “forward” AWGN channel $\hat{\mathbf{x}}'[t] = \mathbf{x}[t] + \mathbf{q}[t]$, where $\mathbf{q}[t]$ is the quantization error, independent of $\mathbf{x}[t]$. This is a common abstraction and is realized via entropy-constrained dithered quantization (ECDQ).¹

As required by the problem formulation, the encoder and decoder incur no delay, but both are allowed to access their entire causal history and adapt to inputs. The encoder’s source-tracking signal $\hat{\mathbf{s}}'[t]$ is produced by the encoder’s K -th order causal linear predictor operating on the encoder’s past reconstructions: $\hat{\mathbf{s}}'[t] = \hat{\mathbf{x}}'[t] + \sum_{k=1}^K \beta_k \hat{\mathbf{s}}'[t - k]$. The decoder uses an L -th order causal linear predictor to reconstruct the current sample $\hat{\mathbf{s}}[t]$: $\hat{\mathbf{s}}[t] = \hat{\mathbf{x}}[t] + \sum_{l=1}^L \gamma_l \hat{\mathbf{s}}[t - l]$. Note that a “standard” (fully predictive) DPCM system for the $\text{AR}(P)$ source uses the P -th order filter, with tap weights $\alpha_1, \dots, \alpha_P$ equal to the source parameters, in both the encoder and decoder predictors. Alternatively, placing all-zero filters in the encoder and decoder predictors makes the system equivalent to standard PCM with no prediction. Therefore, the system here subsumes both DPCM and PCM and admits a type of continuous hybridization between full-prediction and no-prediction systems.

The following standard “DPCM identity” [3] holds:

$$\hat{\mathbf{s}}'[t] - \mathbf{s}[t] = \hat{\mathbf{x}}'[t] - \mathbf{x}[t] = \mathbf{q}[t]. \quad (3)$$

In the following sections, we are interested in the regime of the combination of the following limits:

1. Large spatial dimension: $N \rightarrow \infty$. This allows the notational convenience of analyzing the scalar version of the problem, while still letting the quantizer \mathcal{Q} achieve

¹Additional gain factors make \mathcal{Q} an optimal quantizer for general rates, but they are not needed in high-rate results. See [1] for examples and further references.

vector quantization performance if it codes a large number of independent samples. If $x[t]$ is zero-mean Gaussian with variance Σ_x , the quantization error has variance

$$\Sigma_q = \Sigma_x / (2^{2R} - 1) \quad (4)$$

in the limit. In practice, at finite N , we can replace Σ_x by a slightly larger term, where the maximal loss even for scalar transmission is the equivalent of $\frac{1}{2} \log(\frac{2\pi e}{12}) \approx 0.25$ bits per scalar sample.

2. Sparse erasures: $\hat{s}'[t_\epsilon - 1] = \hat{s}[t_\epsilon - 1]$ before an erasure at t_ϵ . This allows analyzing each erasure separately as their distortion effects interact additively. This happens when erasures occur far apart compared to system memory.
3. High rate: $R \rightarrow \infty$. Consequently, $\Sigma_q \ll \Sigma_z \leq \Sigma_s$, and \mathcal{Q} as described becomes an optimal quantizer.

3.2 One-tap prediction for AR(1) Gaussian sources

We defer the main results on general sources to Section 4. In this section, we show how one-tap DPCM and PCM perform without and with erasures for an AR(1) Gaussian source $s[t]$. We then propose to achieve better performance by using one-tap “leaky prediction,” introduced here not as a novel scheme but to give intuition for the main results.

3.2.1 One-tap DPCM and PCM

Under standard DPCM, the predictors “match” the source and only use one tap of weight, $\gamma_1 = \beta_1 = \alpha_1$. Under standard PCM, $\gamma_1 = \beta_1 = 0$. Without erasures, the distortion is from the quantization error $q[t]$, and is governed by the input variance Σ_x , as in (4). The zero-erasure distortions are:

$$\begin{aligned} D_{\text{DPCM}}(R; 0) &= \frac{\Sigma_z + \alpha_1^2 \Sigma_q}{2^{2R} - 1} = \frac{1}{2^{2R} - 1 - \alpha_1^2} \\ D_{\text{PCM}}(R; 0) &= \frac{\Sigma_s}{2^{2R} - 1} = \frac{1}{(1 - \alpha_1^2)(2^{2R} - 1)} \end{aligned}$$

If erasures occur, the choice of decoder for a given encoder is less obvious. We can let $\hat{x}[t_\epsilon] = 0$ whenever the packet at t_ϵ is erased, in which case the encoder-decoder asynchrony replaces the quantization error $q[t_\epsilon]$ by an error equal to $-x[t_\epsilon]$, which filters through the decoder predictor loop as error propagation for $t \geq t_\epsilon$.

Under DPCM (still with $\gamma_1 = \beta_1 = \alpha_1$), the distortion is

$$D_{\text{DPCM}}(R; \epsilon) = D_{\text{DPCM}}(R; 0) + \epsilon \left[\frac{1}{1 - \alpha_1^2} - \frac{1 - 3\alpha_1^2}{1 - \alpha_1^2} D_{\text{DPCM}}(R; 0) \right]. \quad (5)$$

Under PCM, for the sample at t_ϵ , the decoder can do significantly better by switching from the standard decoder ($\gamma_1 = \beta_1 = 0$) to the DPCM decoder ($\gamma_1 = \alpha_1$). When we refer to PCM throughout the remainder of the paper where erasures occur, we will always mean PCM with this decoder modification during erasures. Therefore,

$$D_{\text{PCM}}(R; \epsilon) = D_{\text{PCM}}(R; 0) + \epsilon [1 - (1 - \alpha_1^2) D_{\text{PCM}}(R; 0)]. \quad (6)$$

In the high-rate limit, the performances in terms of excess distortion ratio [cf. (2)] corresponding to (5) and (6) respectively become:

$$L_{\text{DPCM}}^{\infty}(\lambda) = 1 + \lambda \Sigma_s = 1 + \frac{\lambda}{1 - \alpha_1^2} \quad (7)$$

$$L_{\text{PCM}}^{\infty}(\lambda) = \Sigma_s + \lambda = \frac{1}{1 - \alpha_1^2} + \lambda \quad (8)$$

As anticipated, there is a *critical error threshold*, $\lambda = 1$, such that $L_{\text{DPCM}}^{\infty}(\lambda) < L_{\text{PCM}}^{\infty}(\lambda)$ if $\lambda < 1$ but $L_{\text{DPCM}}^{\infty}(\lambda) > L_{\text{PCM}}^{\infty}(\lambda)$ if $\lambda > 1$.

3.2.2 One-tap leaky prediction

The fact that neither PCM nor DPCM dominates over all error severities suggests that an intermediate encoder predictor, coupled with a suitable decoder predictor, can improve performance over both schemes. A reasonable possibility we consider in this section is to optimize over all one-tap predictors β_1 between the extremes of DPCM ($\beta_1 = \alpha_1$) and PCM ($\beta_1 = 0$). The resulting system is sometimes termed “leaky prediction” [6]. As in the previous section, we use a decoder matched to the encoder predictor ($\gamma_1 = \beta_1$) when there are no erasures, and a DPCM predictor ($\gamma_1 = \alpha_1$) with $\hat{x}[t_\epsilon] = 0$ whenever there is an erasure at time t_ϵ .²

Without erasures, the distortion is again governed by Σ_x through (4). With $x[t] = (\alpha_1 - \beta_1)s[t - 1] - \beta_1q[t - 1] + z[t]$, and the summands being mutually independent, the distortion amounts to:

$$D_{\text{leaky}}(R; 0) = \left[1 + \frac{(\alpha_1 - \beta_1)^2}{1 - \alpha_1^2} \right] (2^{2R} - 1 - \beta_1^2)^{-1}.$$

When there is an erasure at time t_ϵ , the encoder-decoder asynchrony propagates error for $t \geq t_\epsilon$, and replaces the erasure-free identity $\hat{s}[t] - s[t] = q[t]$ by $\hat{s}[t] - s[t] = q[t] - \beta^{(t-t_\epsilon)}(z[t_\epsilon] - \alpha_1q[t_\epsilon - 1] + q[t_\epsilon])$. Taking these errors into account, the distortion is

$$D_{\text{leaky}}(R; \epsilon) = D_{\text{leaky}}(R; 0) + \epsilon \left[\frac{1}{1 - \beta_1^2} - \frac{1 - \alpha_1^2 - 2\beta_1^2}{1 - \beta^2} D_{\text{leaky}}(R; 0) \right]. \quad (9)$$

Fig. 2 shows that the optimal amount of prediction “leakage,” as captured by $\beta_1 = \beta^*$, shifts from better prediction (DPCM, $\beta^* \rightarrow \alpha_1$) to better error resilience (PCM, $\beta^* \rightarrow 0$) as the error severity ϵ increases.

In the high-rate limit, the excess distortion ratio [cf. (2)] corresponding to (9) is

$$L_{\text{leaky}}^{\infty}(\lambda) = \left[1 + \frac{(\alpha_1 - \beta_1)^2}{1 - \alpha_1^2} \right] + \frac{\lambda}{1 - \beta_1^2}. \quad (10)$$

The three expressions (7), (8), and (10) all show distortion composed of two terms, the first being distortion caused by quantization, and the second, distortion caused by channel erasures. In DPCM, the first term is smaller than in PCM, at the expense of the second being larger, while in leaky prediction, the encoder predictor β_1 optimally adjusts the relative weights of the two terms for a given λ . Fig. 3 shows leaky prediction outperforming both DPCM and PCM.

²Or, as an alternative interpretation, we use the erasure-free decoder at an erased sample, but “simulate” the missing decoder input as $\hat{x}[t_\epsilon] = (\alpha_1 - \beta_1)\hat{s}[t_\epsilon - 1]$.

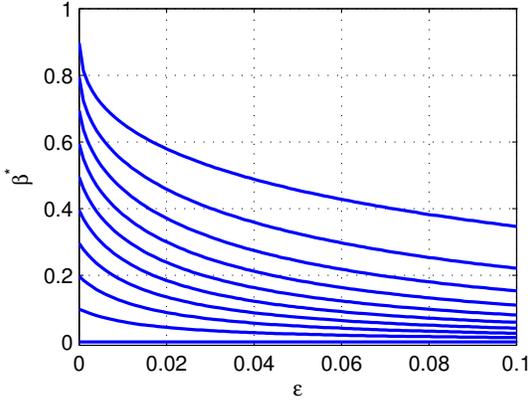


Figure 2: Encoder predictor $\beta_1 = \beta^*$ that minimizes $D_{\text{leaky}}(R; \epsilon)$ for $R = 3$ bits/sample and various values of $\alpha_1 = \{0, 0.1, 0.2, \dots, 0.9\}$ (bottom to top).

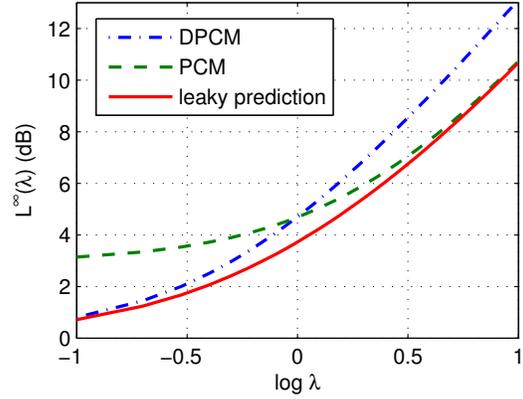


Figure 3: DPCM, PCM, and one-tap leaky prediction performance compared for an AR(1) source with $\alpha_1 = 0.7$, $R = 3$. The critical error threshold is at $\lambda = 1$.

4 Main results

The previous section shows that for AR(1) sources, one-tap leaky prediction performs better than DPCM and PCM. As we will see shortly, one-tap predictors are not optimal for AR(1) sources when there are erasures, even in the high-rate limit. In this section, we derive optimized predictors for the linear predictive quantization system for colored Gaussian sources in the high-rate limit.

In order to be able to evaluate the systems we develop more meaningfully, we first develop a simple lower bound on the distortion achievable by any causal system.

4.1 Performance bound

As before, $D_s(R)$ is the distortion-rate function of the source $s[t]$. $D_\Delta(R)$ is the expected additional distortion caused by each erasure under some particular scheme, and D_ϵ is the expected distortion on just the erased sample. The distortion can be lower-bounded as:

$$\begin{aligned} D(R; \epsilon) &= (1 - \epsilon)D(R; 0) + \epsilon D(R; 0) + \epsilon D_\Delta(R) \\ &\geq (1 - \epsilon)D_s(R) + \epsilon D_\epsilon \\ &\geq (1 - \epsilon)2^{-2R} + \epsilon \end{aligned}$$

$D(R; 0) \geq D_s(R)$ by the definition of the distortion-rate function. $D(R; 0) + D_\Delta(R)$ is the total distortion on an erasure at some time t_ϵ , plus additional distortion thereafter. This is greater than the distortion at t_ϵ only, which is in turn lower-bounded by the error variance of optimal prediction of $s[t_\epsilon]$ from its entire causal past $s[t_\epsilon - 1], s[t_\epsilon - 2], \dots$. This is simply the variance of the innovation process ($\Sigma_z = 1$).

Consequently, in the high-rate limit, the excess distortion ratio [cf. (2)] satisfies

$$L^\infty(\lambda) \geq 1 + \lambda. \quad (11)$$

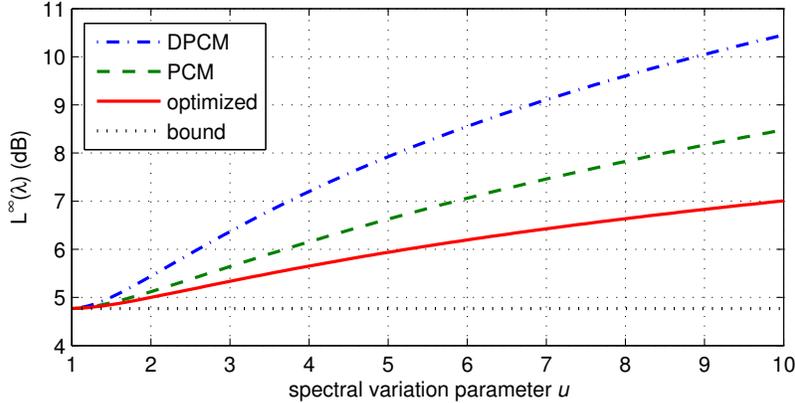


Figure 4: Performance comparison ($\lambda = 2$) between three systems and the bound of (11) as a function of the spectral variation parameter u , for the source with discontinuous spectrum of Example 3. The performance gaps from either DPCM or PCM to the system with optimized parameters are unbounded when u is taken arbitrarily large.

4.2 Achievable performance for colored sources at high rate

We have the next result on the performance achievable by schemes of the form of Fig. 1.

Theorem 1. *Let $s[t]$ be a stationary Gaussian source with power spectral density $\Phi_s(f)$ and unit entropy power, i.e., $N_s = \exp \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \Phi_s(f) df = 1$. The excess distortion ratio [cf. (2)]*

$$L^\infty(\lambda) = \min_{\Phi_v(f): N_v=1} \left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\Phi_s(f)}{\Phi_v(f)} df + \lambda \int_{-\frac{1}{2}}^{\frac{1}{2}} \Phi_v(f) df \right\} \quad (12)$$

is achievable, where $\Phi_v(f)$ is any valid power spectrum and N_v is its entropy power.

We call the minimizing $\Phi_v^*(f)$ the spectrum of a *virtual source*. The optimal encoder predictor to achieve this $L^\infty(\lambda)$ is the DPCM predictor for this virtual source, rather than for the actual source $s[t]$. We will also see that this encoder predictor is a spectral compromise between the DPCM predictor and PCM predictor for $s[t]$.

It is easily verified that $L^\infty(\lambda = 1)$ is achieved by $\Phi_v^*(f) = \sqrt{\Phi_s(f)}$.

Example 2. Applying the $\lambda = 1$ case to an AR(1) Gaussian source, the virtual source spectrum $\sqrt{\Phi_s(f)} = 1/|1 - \alpha_1 e^{-j2\pi f}|$ is clearly not AR(1). Thus, the possibility of a one-tap optimal encoder predictor for $s[t]$ is precluded.

Example 3. Applying the result beyond finite-order AR(P) sources, let $s[t]$ be a colored Gaussian source with a two-level spectrum: $\Phi_s(f) = u > 0$ for $|f| \leq 1/4$ and $\Phi_s(f) = u^{-1}$ for $1/4 < |f| \leq 1/2$. (Note that with this parameterization, $N_s = 1$ for all $u > 0$.) The optimal virtual source spectrum will also be two-level, i.e., $\Phi_v^*(f) = v^*$ for $|f| \leq 1/4$ and $\Phi_v^*(f) = v^{*-1}$ for $1/4 < |f| \leq 1/2$. The minimization (12) then amounts to $\min_{v>0} u/v + v/u + \lambda v + \lambda/v$. Solving gives $v^* = \sqrt{(\lambda + u)/(\lambda + u^{-1})}$ and $L^\infty(\lambda) = \sqrt{(\lambda + u)(\lambda + u^{-1})}$.

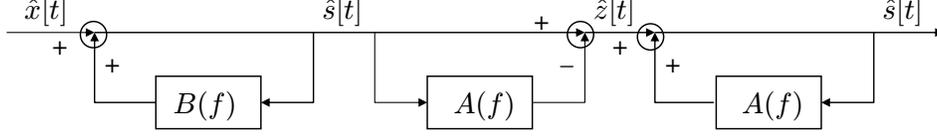


Figure 5: The equivalent system at the decoder. When a sample is not erased, the entire decoder implements $1/(1 - B(f))$. When a sample is erased, the first two filters are effectively bypassed as the decoder sets $\hat{z}[t]$ to 0.

Compare with $L_{\text{DPCM}}^\infty(\lambda) = 1 + \lambda(u + u^{-1})/2$ and $L_{\text{PCM}}^\infty(\lambda) = (u + u^{-1})/2 + \lambda$ for this source. For $u \gg 1$, $L^\infty(\lambda)$ grows as \sqrt{u} , while $L_{\text{DPCM}}^\infty(\lambda)$ and $L_{\text{PCM}}^\infty(\lambda)$ both grow as u , so the performance gap, as measured by excess distortion incurred by DPCM or PCM over the proposed system, can be arbitrarily large. Fig. 4 compares these systems and the performance bound.

4.3 Analysis: Constructive derivation of Theorem 1

Let $B(f) = \sum_{k=1}^K \beta_k e^{-j2\pi fk}$, $\Gamma(f) = \sum_{l=1}^L \gamma_l e^{-j2\pi fl}$ be respectively the frequency responses of the encoder and decoder predictors. Let $\{\cdot\}_*$ denote the causal minimum-phase filter with spectrum in the argument. We show next that a linear predictive quantization system, with encoder predictor $B^*(f) = 1 - \{1/\Phi_v^*(f)\}_*$, decoder predictor $\Gamma^*(f) = B^*(f)$ on a non-erased sample, and decoder predictor $\Gamma^*(f) = 1 - \{1/\Phi_s(f)\}_*$ on an erased sample, achieves $L^\infty(\lambda)$ in Theorem 1.

Without erasures,

$$\Sigma_{s-\hat{s}} = \left| 1 - \frac{1 - B(f)}{1 - \Gamma(f)} \right|^2 \Sigma_s + \left| \frac{1 - B(f)}{1 - \Gamma(f)} \right|^2 \frac{\Sigma_x}{2^{2R} - 1},$$

so in the high-rate limit, $\Sigma_{s-\hat{s}}$ is minimized by choosing a matched decoder predictor $\Gamma(f) = B(f)$, leaving the erasure-free distortion as

$$D_*(R; 0) = \Sigma_x 2^{-2R}, \quad (13)$$

where $\Sigma_x = \int_{-\frac{1}{2}}^{\frac{1}{2}} \Phi_x(f) df = \int_{-\frac{1}{2}}^{\frac{1}{2}} |1 - B(f)|^2 \Phi_s(f) df$. The minimal $D_*(R; 0)$ can be achieved by choosing the unique stable, strictly causal predictor $B(f)$ that minimizes Σ_x . This occurs when $\Phi_x(f)$ is white, hence $|1 - B(f)|^2 = N_s/\Phi_s(f)$ and the optimal erasure-free encoder predictor is $B(f) = 1 - \{N_s/\Phi_s(f)\}_*$. For example, an AR(P) source $\{s[t]\}_t$ with innovation variance $\Sigma_z = 1$ has power spectral density $\Phi_s(f) = 1/|1 - A(f)|^2$ from Section 2.1. The optimal erasure-free encoder predictor $B(f)$ is then the standard source-matching DPCM predictor $B(f) = A(f)$, as expected.

When erasures are possible, the encoder predictor $B(f)$ is time-invariant as the encoder receives no feedback. However, the decoder can adapt. Referring to Fig. 5, suppose the decoder internally derives an estimated innovation sequence $\{\hat{z}[t]\}_t$ by filtering $\{\hat{x}[t]\}_t$ through $1/(1 - B(f))$ followed by $1 - A(f)$. The decoder is to produce the best estimate of $s[t]$ by applying a third filter, which we claim to be $1/(1 - A(f))$. In the high-rate limit, $\hat{z}[t] \rightarrow z[t]$. An erased sample of $\hat{x}[t]$ corresponds to a missing sample of $\hat{z}[t]$, which the decoder ‘‘patches’’ with its best causal estimate $\mathbf{E}\{\hat{z}[t|t-1, t-2, \dots]\} = \mathbf{E}\{z[t]\} = 0$. It

is clear that the patched innovation sequence $\hat{z}[t]$ contains all the information the decoder has about $s[t]$ and applying the reconstruction filter $1/(1 - A(f))$ to $\hat{z}[t]$ reconstructs the best estimate of $s[t]$. This means that, for unerased samples, the decoder's effective decoder predictor is $\Gamma(f) = B(f)$, while for erased samples, it is $\Gamma(f) = A(f)$. Furthermore, this adds an error propagation process that is $\hat{z}[t_\epsilon]\delta[t - t_\epsilon]$ filtered by $1/(1 - B(f))$, whenever $\hat{x}[t_\epsilon]$ is an erased sample. In the high-rate limit, this results in an additional distortion of $D_\Delta = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1/|1 - B(f)|^2 df$ for each erasure. Thus, the average distortion with erasures for this setup at high rate is

$$\begin{aligned} D_*(R; \epsilon) &= D_*(R; 0) + \epsilon D_\Delta \\ &= 2^{-2R} \int_{-\frac{1}{2}}^{\frac{1}{2}} |1 - B(f)|^2 \Phi_s(f) df + \epsilon \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{|1 - B(f)|^2} df. \end{aligned} \quad (14)$$

Substituting in $\lambda = \epsilon/2^{-2R}$ [cf. (2)] gives immediately

$$\min_{\Phi_v(f): N_v=1} L_*^\infty(\lambda) = \min_{\Phi_v(f): N_v=1} \left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\Phi_s(f)}{\Phi_v(f)} df + \lambda \int_{-\frac{1}{2}}^{\frac{1}{2}} \Phi_v(f) df \right\},$$

thereby proving Theorem 1.

As in Section 3.2, there are two terms in (12) of Theorem 1, in which the first is caused by quantization and the second by channel erasures. The choice of encoder predictor $B(f)$ again weights the two terms. We can interpret $\Phi_v(f) = 1/|1 - B(f)|^2$ as the power spectral density of the virtual source $v[t]$ against which $B(f)$ is trying to predict. Depending on λ , the optimizing virtual source is either spectrally more like $s[t]$ (smaller λ) or more white (larger λ). If the virtual source spectrum matches $\Phi_s(f)$ of the real source, then $B(f)$ implements DPCM and $L_*^\infty(\lambda) = 1 + \lambda \Sigma_v$. If the virtual source spectrum is white, then $B(f)$ implements PCM and $L_*^\infty(\lambda) = \Sigma_v + \lambda$. The minimizing virtual source has spectrum between these, corresponding to $B(f)$ implementing a hybrid between DPCM and PCM.

References

- [1] R. Zamir, Y. Kochman and U. Erez. "Achieving the Gaussian Rate-Distortion Function by Prediction." *IEEE Trans. Info. Theory*, vol. 54, no. 7, pp. 3354-3364, July, 2008.
- [2] N. Ma and P. Ishwar. "On Delayed Sequential Coding of Correlated Sources." Submitted to *IEEE Trans. Info. Theory*, 2008. Available at <http://arxiv.org/abs/cs/0701197>.
- [3] N. S. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice-Hall. Englewood Cliffs, NJ, 1984.
- [4] K. Chang and R. Donaldson. "Analysis, Optimization, and Sensitivity Study of Differential PCM Systems Operating on Noisy Communication Channels." *IEEE Trans. Communications*, vol. 20, no. 3, pp. 338-350, Feb., 1972.
- [5] T. Berger. *Rate Distortion Theory, A Mathematical Basis for Data Compression*. Prentice-Hall. Englewood Cliffs, NJ, 1971.
- [6] H. C. Huang, W. H. Peng, T. Chiang. "Advances in the Scalable Amendment of H.264/AVC." *IEEE Comm. Magazine*, vol. 45, no. 1, pp. 68-76, Jan., 2007.