Causal transmission of colored source frames over a packet erasure channel

Ying-zong Huang, Yuval Kochman & Gregory Wornell
Massachusetts Institute of Technology
Dept. of Electrical Engineering and Computer Science
Cambridge, MA 02139
{zong,yuvalko,gww}@mit.edu

Abstract

Motivated by the streaming video application, 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 DPCM performs well without erasures and PCM performs well with many erasures, but not vice versa, we show that the proposed solution improves performance over them 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 for its intuitive appeal and simplicity.

Classically, scalar DPCM was used for audio and scanline image and video coding, due to its lower bit rate relative to, e.g., PCM and the practicality of 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 instances. 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. In [2], an implicit consequence of the analysis applied 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 errors. In applications such as real-time internet video conferencing, it is realistic 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 determine 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) since the very beginning of DPCM, it was recognized that adverse effects of channel errors were reduced by altering predictors in DPCM [3, 4]; and (2) engineering practice in, e.g., video coding gives the impression that good prediction (interframe coding) can be traded for robustness against errors (intraframe coding).

After we formulate the problem more precisely in Section 2, we consider the linear predictive quantization system that generalizes both DPCM and PCM as a solution 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. Finally an end-to-end distortion is found for that rate. We look for the encoder and decoder pair that optimizes the end-to-end distortion performance in the system. More specifically:

2.1 Source model

Let $\{\mathbf{s}[t]\}_t = \{(s_1[t], ..., s_N[t])\}_t$ be a vector of N parallel source sequences, which can be viewed as a sequence of frames. In deference to the analogy of video coding, 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 $\alpha_1, ..., \alpha_P$ 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,

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

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 the P parameters $\alpha_1, ..., \alpha_P$ (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 NR bits. The channel has two states. The channel state is an i.i.d. process described by an erasure probability ϵ. In the normal or no-erasure state occuring with probability $1-\epsilon$, the input is reproduced exactly at the output. In the erasure state occuring with probability ϵ, the output is a special lost packet symbol $\varnothing$ 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 $\mathbf{s}[t], \mathbf{s}[t-1], ...$ as input and produces a packet $\mathcal{E}_t = \mathcal{E}(\mathbf{s}[t], \mathbf{s}[t-1], ...)$ of NR bits. The causal decoder takes current and past receptions $\mathcal{C}(\mathcal{E}_t), \mathcal{C}(\mathcal{E}_{t-1}), ...$ and outputs $\hat{\mathbf{s}}[t] = \mathcal{D}(\mathcal{C}(\mathcal{E}_t), \mathcal{C}(\mathcal{E}_{t-1}), ...)$, a reproduction of the newest source vector $\mathbf{s}[t]$.

2.4 Performance and objective

The average end-to-end distortion per scalar sample between source $\mathbf{s}[t]$ and its reproduction $\hat{\mathbf{s}}[t]$ at time t is 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:

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

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

$$L(R;\epsilon) \triangleq D(R;\epsilon)/D_s(R) = D(R;\epsilon)/2^{-2R}$$

for the supplied R and ϵ, 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

$$L^{\infty}(\lambda) \triangleq \lim_{R\to\infty} L(R;\lambda 2^{-2R}), \qquad (0 < \lambda < \infty) \tag{2}$$

3 Preliminary insights

It is not known what general causal encoder/decoder pair minimizes average distortion when erasures are possible. However, we can consider the problem within the class of linear predictive quantization systems for reasons of tractability and importance that such systems play in engineering practice. We defer the main results to the next section. In this section, we first show how standard DPCM and PCM perform without, then with, erasures for an AR(1) source. We then propose a way to optimize predictors for this simple source by "leaky prediction," which is introduced here not as a novel scheme but to give intuition for more general sources in Section 4.

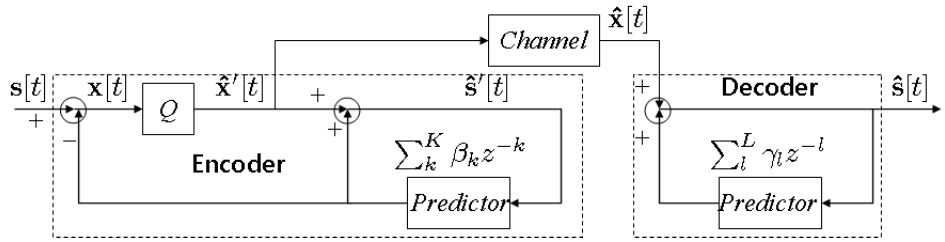

Figure 1: A causal, linear predictive quantization system model. $\mathcal{Q}$ is a quantizer coding the N-vector $\mathbf{x}[t]$.

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. 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.

A "standard" (fully predictive) DPCM system for the AR(P) source uses the P-th order filter, with tap weights $\alpha_1, ..., \alpha_P$ equal to the source parameters, in both the encoder and decoder predictors. On the other hand, placing all-zero filters in the encoder and decoder predictors makes the system equivalent to 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 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 outputs $\hat{\mathbf{x}}'[t] \in \hat{\mathcal{X}}_t \subseteq \mathbf{R}^N$ at rate R per scalar sample ($H(\hat{\mathbf{x}}'[t])/N = R$). We model the quantizer $\mathcal{Q}$ as a "forward" AWGN channel, where $\mathbf{q}[t]$ is independent of $\mathbf{x}[t]$. This is a common abstraction and is realized via entropy-constrained dithered quantization. (See [1] for examples and further references.)

Since the channel introduces erasures of an entire packet only, we neglect bitstream encoding and decoding, by assuming the channel passes $\hat{\mathbf{x}}'[t]$ itself to the decoder as $\hat{\mathbf{x}}[t]$ in the no-erasure state. The following standard "DPCM identity" [3] holds:

$$\hat{\mathbf{s}}'[t] - \mathbf{s}[t] = \hat{\mathbf{x}}'[t] - \mathbf{x}[t] = \mathbf{q}[t] \tag{3}$$

where $\mathbf{q}[t]$ is the quantization error and $\hat{\mathbf{s}}'[t]$ is the encoder's current reconstruction.

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 $\hat{\mathbf{s}}[t]$: $\hat{\mathbf{s}}[t] = \hat{\mathbf{x}}[t] + \sum_{l=1}^{L} \gamma_l \hat{\mathbf{s}}[t-l]$.

We are interested in the regime of the combination of the following limits:

1. Large spatial dimension: $N \to \infty$. This allows analyzing the scalar version of the problem, while remembering that the quantizer $\mathcal{Q}$ still achieves vector quantization performance, as it codes numerous independent samples. If $x[t]$ is zero-mean Gaussian with variance Σ_x, the quantization error has variance

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

Practically, a finite dimension suffices, and at $N = 1$, results are correct up to a loss of $\frac{1}{2}\log(\frac{2\pi e}{12})$ 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: $R \to \infty$. Consequently, $\Sigma_q \ll \Sigma_z \leq \Sigma_s$.

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

Before more general results, we first interpret some standard systems as special cases of the linear predictive quantization system, and show their performances for coding the relatively simple AR(1) Gaussian source. Then we find their performances when there are channel erasures. Finally, we show ways to improve upon them via "leaky prediction."

3.2.1 DPCM and PCM without error

For the standard DPCM system, the predictors are "matched" to the source and use only one tap of weight $\beta_1 = \gamma_1 = \alpha_1$. In this case, the differential signal $x[t]$ is Gaussian with variance $\Sigma_z + \alpha_1^2 \Sigma_q$, and so

$$D_{\mathrm{DPCM}}(R;0) = \frac{\Sigma_x}{2^{2R} - 1} = \frac{1}{2^{2R} - 1 - \alpha_1^2}$$

Contrast this with the PCM encoder, which achieves distortion-rate

$$D_{\mathrm{PCM}}(R;0) = \frac{\Sigma_s}{2^{2R} - 1} = \frac{1}{(1 - \alpha_1^2)(2^{2R} - 1)}$$

For $R > 1/2$ (all R if the quantizer $\mathcal{Q}$ is optimal), we observe that $D_{DPCM}(R;0) < D_{PCM}(R;0)$ for all $\alpha_1 \neq 0$, as is well known.[1]

3.2.2 DPCM and PCM with error

When an erasure occurs, the encoder-decoder asynchrony replaces the quantization error $q[t]$ by an error which equals $-x[t]$ instead, which then gets filtered through the decoder predictor loop. With an imperfect channel, the choice of decoder for a given encoder becomes less obvious.

Under DPCM, where we still use $\beta_1 = \gamma_1 = \alpha_1$, the distortion-rate function is

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

For PCM, $\beta_1 = \gamma_1 = 0$ is standard. However, for the sample on which an erasure occurs, the decoder can do significantly better by switching to the DPCM decoder ($\gamma_1 = \alpha_1$). When we refer to PCM throughout the remainder of the paper, we will always mean PCM with this decoder modification during erasures. Therefore,

$$D_{\mathrm{PCM}}(R;\epsilon) = D_{\mathrm{PCM}}(R;0) + \epsilon \left[1 - (1 - \alpha_1^2) D_{\mathrm{PCM}}(R;0) \right]$$

It can be verified that there is some critical error threshold $\epsilon_0(R)$ such that $D_{\mathrm{DPCM}}(R;\epsilon) < D_{\mathrm{PCM}}(R;\epsilon)$ if $\epsilon < \epsilon_0$ but $D_{\mathrm{DPCM}}(R;\epsilon) > D_{\mathrm{PCM}}(R;\epsilon)$ if $\epsilon > \epsilon_0$.

[1] It is also known that a single-tap, source-matching predictor is not optimal for the no-erasure case except in the high-rate limit [6, 7].

3.2.3 One-tap leaky prediction

The fact that neither PCM nor DPCM dominates over all error severities suggests that a more flexible predictor in the encoder, coupled with a suitable predictor in the decoder, can improve performance over both schemes.

In lieu of the extremal encoder predictors of DPCM ($\beta_1 = \alpha_1$) or PCM ($\beta_1 = 0$), we optimize over all one-tap predictors β_1. The result is sometimes termed "leaky prediction" [8], as some of the original signal energy in $s[t]$ leaks through to the next time step. Unlike DPCM or PCM, there is no standard "leaky prediction" system as a convention to draw on for the decoder predictor. Nevertheless, analogous to the DPCM or PCM cases, we use a decoder matched to the encoder predictor (i.e., $\gamma_1 = \beta_1$) when there are no erasures, and a DPCM predictor (i.e., $\gamma_1 = \alpha_1$) whenever there is an erasure. [2]

Without erasures, the error signal $s[t]-\hat{s}[t]$ has variance $D_{\text{leaky}}(R;0) = \Sigma_q$. The quantizer input is $x[t] = (\alpha_1-\beta_1)s[t-1]-\beta_1 q[t-1]+z[t]$. As the summands are mutually independent terms, 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 no-erasure identity $\hat{s}[t]-s[t] = q[t]$ by $\hat{s}[t]-s[t] = q[t]-\beta^{(t-t_\epsilon)}(z[t_\epsilon]-\alpha_1 q[t_\epsilon - 1] + q[t_\epsilon])$. Taking these errors into account, the distortion rate function is

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

Fig. 2 shows that the optimal amount of prediction "leakage" as captured by $\beta_1 = \beta^*$ produces a hybrid between DPCM and PCM, and shifts from better prediction (DPCM, $\beta^* \to \alpha_1$) to better error resilience (PCM, $\beta^* \to 0$) as the error severity ϵ increases. Fig. 3 shows leaky prediction outperforming both DPCM and PCM.

3.2.4 High-rate limit

In the high-rate limit, the performances in terms of excess distortion ratio (2) become:

$$L^{\infty}_{\text{DPCM}}(\lambda) = 1 + \lambda\Sigma_s = 1 + \frac{\lambda}{1-\alpha_1^2} \tag{5}$$

$$L^{\infty}_{\text{PCM}}(\lambda) = \Sigma_s + \lambda = \frac{1}{1-\alpha_1^2} + \lambda \tag{6}$$

$$L^{\infty}_{\text{leaky}}(\lambda) = \left[1 + \frac{(\alpha_1 - \beta_1)^2}{1-\alpha_1^2}\right] + \frac{\lambda}{1-\beta_1^2} \tag{7}$$

All three expressions 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, as expected. In the high-rate limit, the threshold for PCM being better than DPCM is met with $\lambda = 1$, i.e., $\epsilon_0(R) = 2^{-2R}$. However, in leaky prediction, the encoder predictor β_1 adjusts the weight between the two terms, and can be optimized for a given λ, therefore outperforming both PCM and DPCM.

[2] Or, equivalently, when an erasure occurs, we use the error-free decoder but "simulate" the missing decoder input as $\hat{x}[t] = (\alpha_1 - \beta_1)\hat{s}[t-1]$.

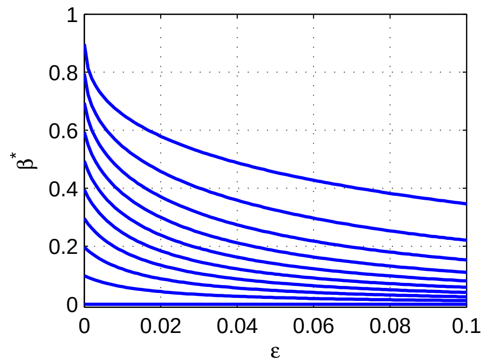

Figure 2: Encoder predictor $\beta_1 = \beta^*$ that minimizes $D_{\text{leaky}}(R;\epsilon)$ for $R = 3$ bits/sample and various values of $\alpha_1 = \{0, 0.1, 0.2, ..., 0.9\}$ (bottom to top).

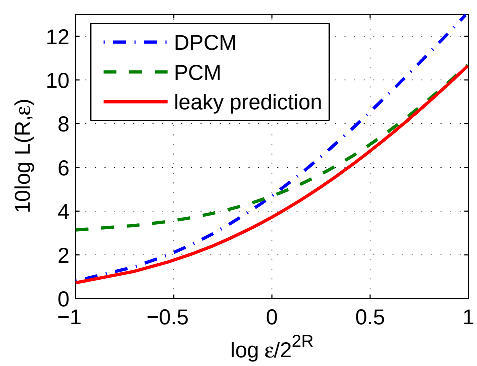

Figure 3: DPCM, PCM, and one-tap leaky prediction performance compared for an AR(1) source with $\alpha_1 = 0.7$, $R = 3$. The critical error threshold is nearly $\epsilon_0(R) = 2^{-2R}$.

4 Main results

The previous section shows that for AR(1) sources, one-tap leaky prediction gives better performance 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 performance bound that no causal system can outperform.

4.1 Performance bound

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

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

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

In the high-rate limit, the excess distortion ratio satisfies

$$L^\infty(\lambda) \geq 1 + \lambda \tag{8}$$

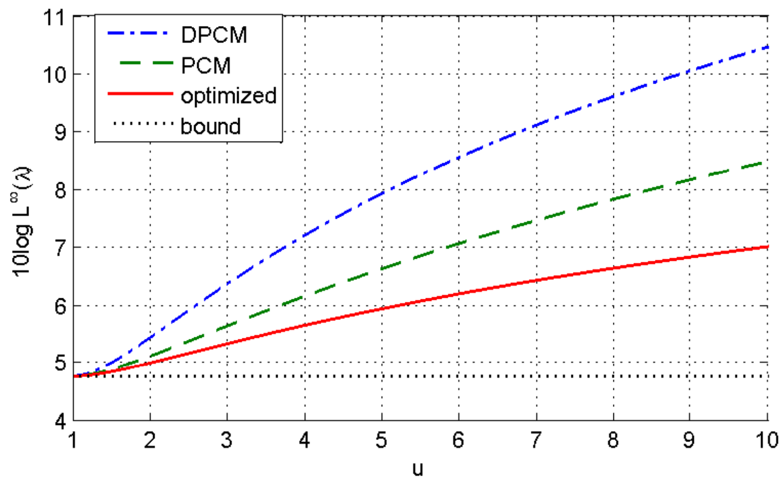

Figure 4: Performance comparison ($\lambda = 2$) between three systems and the bound of (8), for sources in Example 3. The performance gaps from both DPCM or PCM to the system with optimized parameters is unbounded when u is taken arbitrarily large.

4.2 Achievable performance for colored sources at high rate

We have the following 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 (2) of*

$$L^\infty(\lambda) = \min_{\Phi_v(f):N_v=1} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\Phi_s(f)}{\Phi_v(f)} df + \lambda \int_{-\frac{1}{2}}^{\frac{1}{2}} \Phi_v(f) df \tag{9}$$

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, not 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 $\sqrt{\Phi_s(f)} = 1/\left|1 - \alpha_1 e^{-j2\pi f}\right|$ 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 these 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 (9) then amounts to $\min_{v>0} u/v + v/u + \lambda v + \lambda/v$. Solving gives $v^* = \sqrt{(\lambda + u)/(\lambda + u^{-1})}$.

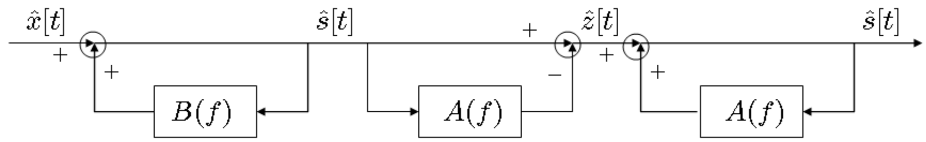

Figure 5: The equivalent system at the decoder. When a sample is not erased, the entire decoder implements $1/(1-B(f))$. When a sample is erased, the first two filters are effectively bypassed as the decoder sets $\hat{z}[t]$ to 0.

Compare with $L_{\text{DPCM}}(\lambda) = 1 + \lambda\left(u + u^{-1}\right)/2$ and $L_{\text{PCM}}(\lambda) = (u + u^{-1})/2 + \lambda$ for this source. For $u \gg 1$, $L_*(\lambda)$ grows as $\sqrt{u}$, while $L_{\text{DPCM}}(\lambda)$ and $L_{\text{PCM}}(\lambda)$ 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. 3 compares these systems and the performance bound.

4.3 Achieving the performance by predictive schemes

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-domain representations 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,

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

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

$$D_*(R; 0) = \Sigma_x 2^{-2R} \tag{10}$$

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 error-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 error-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, $\hat{z}[t] \to z[t]$. 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 $\mathbf{E}\hat{z}[t|t-1, t-2, ...] = \mathbf{E}z[t] = 0$. 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 $D_\Delta = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1/\left|1-B(f)\right|^2 df$ for each erasure. Thus, the average distortion with erasures for this setup is

$$D_*(R;\epsilon) = D_*(R;0) + \epsilon D_\Delta = 2^{-2R} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left|1-B(f)\right|^2 \Phi_s(f) df + \epsilon \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\left|1-B(f)\right|^2} df \quad (11)$$

Substituting in $\lambda = \epsilon/2^{-2R}$ as previously defined gives immediately

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

thereby proving Theorem 1. Analogous to (5), (6), there are two terms in (9), 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 $\Phi_v(f) = 1/\left|1-B(f)\right|^2$ as the power spectral density of some *virtual source* $v[t]$ against which $B(f)$ is trying to predict. Depending on λ, which is the severity of erasures relative to the error-free quantization error, 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." *Proc. Int. Symp. Info. Theory (ISIT)*, pp. 803-807. Seattle, July, 2006.

[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] A. N. Kim and T. A. Ramstad. "Improving the Rate-Distortion Performance of DPCM." *Proc. Int. Symp. Signal Processing, Appl.*, pp. 97-100. Paris, July, 2003.

[7] M. Naraghi-Pour and D. L. Neuhoff. "Mismatched DPCM Encoding of Autoregressive Processes." *IEEE Trans. Info. Theory*, vol. 36, no. 2, pp. 296-304, Mar., 1990.

[8] 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.