ON DECODER-LATENCY VERSUS PERFORMANCE TRADEOFFS IN DIFFERENTIAL PREDICTIVE CODING

Prakash Ishwar and Kannan Ramchandran

Department of Electrical Engineering and Computer Sciences,
University of California, Berkeley, CA 94720, USA.
{ishwar,kannanr}@eecs.berkeley.edu

ABSTRACT

Theoretical analysis of differential predictive coding (DPC) has almost exclusively focused on scalar quantizers and the high-rate regime for tractability reasons. As a result, the role of non-causal decoding in improving the quality has been largely ignored in the literature. In this work we conduct a rigorous performance analysis of DPC-based schemes under a simple independent, vector-Gaussian, AR-1 source model and large-block (as opposed to high-rate) asymptotics. This analysis reveals that non-causal decoding can offer a significant relative improvement in the mean squared error (by as much as 3 dB) at medium to low rates (0.1–0.5 bit per sample) for sources having strong temporal correlation. Furthermore, most of this relative improvement can be attained with a modest decoder-latency. At very high and very low rates, the gains are negligible.

1. INTRODUCTION

Differential predictive coded modulation (DPCM) is a popular and well-established predictive compression method with a long history of development (see [1, 2] and the references therein). DPCM has had wide impact in the evolution of compression standards for speech, image, audio, and video coding. The classical DPCM system consists of a causal predictive encoder and a causal decoder. This is in line with applications having low delay/latency tolerance at both encoder and decoder. However, there are many interesting cases where these constraints can be relaxed. There are three additional scenarios corresponding to this: (i) non-causal encoder and decoder; (ii) non-causal encoder and causal decoder; and (iii) causal encoder and non-causal decoder. Application examples of these include, respectively, non-real-time display of non-real-time video for (i), zero-latency display of non-real-time encoded video for (ii), and non-real-time display of live video for (iii). This paper addresses the last scenario. We are motivated to ask the following question: how much performance gain can be had as a function of the decoding-latency?

Much of the theoretical analysis of DPCM has focused on scalar quantizers and the high-rate regime for tractability reasons. As a result of the high-rate assumption, as will become clear in the sequel, the potential performance gains due to non-causal decoding[1] of causally encoded DPCM data (case (iii) above) become negligible. In this work, we are motivated to quantify the gains in scenario (iii) by relaxing the high-rate assumptions. However, for ease of analytical tractability in this case, we resort to a vector-DPCM analysis. Specifically, the scalar source is replaced by a large-block-length vector source, which we analyze for a simple vector-Gaussian AR-1 model, as described in the next section[2]. Our work here has some overlap with that of [2] which addresses numerically optimized R-D optimal filters for pre- and post-filtering based on scalar quantizers. Their focus was on the design of these filters for improving the performance of classical DPCM. Here, our focus is on obtaining a closed-form quantitative characterization of the tradeoff between performance and decoder-latency. Our analysis reveals that temporal smoothing at the decoder offers significant gains in quality (as much as 3 dB) at medium to low rates for sources with strong temporal correlation. Furthermore, most of these gains can be attained with a modest smoothing-latency.

2. MODELING ASSUMPTIONS

2.1. Source model

Let $\{\mathbf{X}_n\}_{n=-\infty}^{n=\infty}$ be a real, zero-mean, stationary, vector-Gaussian, process[3]. The components (of the vector-process) are independently evolving scalar AR-1 processes [3] and are identically distributed[4]. We further assume that the block-length of the vectors is infinite to facilitate the application of results from rate-distortion (R-D) theory [4] which are based on large block-length asymptotics. However, we allow the encoding rate to be arbitrary. The approach followed here is almost contrary to conventional theoretical analyses of DPC schemes where the block-length of the source vectors typically equals one (scalar DPC) but the encoding rate is assumed to be high to make the analysis tractable. Specifically,

$$\mathbf{X}_n = \rho \mathbf{X}_n + \mathbf{W}_{n-1}, \quad \forall n \in \mathbb{Z},$$

where $|\rho| \in (0,1)$, $\mathbf{X}_n \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \mathrm{I}_\infty)$ for all $n \in \mathbb{Z}$, and $\mathbf{W}_n \sim \mathcal{N}(\mathbf{0}, \sigma_W^2 \mathrm{I}_\infty)$ is an i.i.d. vector-Gaussian innovations sequence with $\sigma_W^2 = (1-\rho^2)\sigma_X^2$ for stationarity. Here, $\mathbf{W}_n$ is independent of $\mathbf{X}_{-\infty}^{n-1}$ for all $n \in \mathbb{Z}$ and I_∞ is the infinite identity matrix corresponding to infinite block-length source vectors. As noted earlier, the spatial whiteness assumption on the innovation vector-process is for simplifying the presentation and can be relaxed. However, it would require significant additional work to get rid of the temporal independence assumption on the innovation

[1] Note that we will refer to non-causal decoding throughout the sequel as a "smoothing" operation.

[2] The analysis for scalar quantizers is part of our ongoing work.

[3] We shall abbreviate $\{\mathbf{Z}_n\}_{n=k}^{n=l}$ to $\mathbf{Z}_k^l$

[4] The independence and identically distributed conditions are assumed only for simplifying the exposition and can be relaxed.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

sequence but this is a reasonable assumption to make in a number of applications. We draw motivation from optical-flow models for video coding [5] with $\mathbf{X}_{-\infty}^{+\infty}$ representing the evolution of the video-innovations process *along motion-compensated trajectories* for a group of adjacent pixels. Our model is also relevant to the new class of motion-compensated 3-D wavelet video codecs where motion compensation is done "out of the predictive loop" [6, 7, 8].

2.2. Model for predictive coding

Figure 1 depicts the steady-state *functional* block-diagram of a DPC system consisting of a causal encoder and a possibly non-causal decoder. We are interested in analyzing the R-D (with mean squared error (MSE) as the distortion criterion) performance of such a system particularly with regard to quantifying the gains of non-causal smoothing at the decoder. The objective of this study is to understand the improvement of MSE as a function of the smoothing delay ($m \geq 0$). The MSE for any block $\mathbf{X}_n$ depends on the following parameters: (i) the source variance σ_X^2, (ii) the degree of temporal correlation or memory in the vector-process as measured by the parameter $|\rho|$, (iii) the rate of encoding R, and (iv) the smoothing delay m.

DPC-encoding proceeds by first forming the *causal* minimum mean square error (MMSE) estimate [9] of the current block at time n based on *all* the information available at the decoder until and including time $n-1$. The prediction error $\mathbf{e}_n$ between the current block $\mathbf{X}_n$ and the causal MMSE estimate is compressed to rate R bits per component of the error vector using an *ideal* R-D encoder. The rate-R, R-D reconstruction vector for $\mathbf{e}_n$ is denoted by $\widetilde{\mathbf{e}}_n$. Hence, the causal MMSE estimate is given by $\mathbb{E}[\mathbf{X}_n|\widetilde{e}_{-\infty}^{n-1}]$. Since the source vector-process is Gaussian, the block-length is infinity, and the R-D coder is ideal, in the steady-state, the prediction error vector-process is also Gaussian and stationary. Hence, the ideal R-D quantized codewords for the prediction error vector-process are statistically related to the unquantized prediction error through the rate-R *forward R-D test-channel* depicted in Fig. 1 [4, 10]. Specifically,

$$\widetilde{\mathbf{e}}_n = \alpha\left(\mathbf{e}_n + \mathbf{Q}_n\right), \quad \forall n \in \mathbb{Z},$$

where, $\alpha(R) := 1 - 2^{-2R}$ and $\mathbf{Q}_n \sim \mathcal{N}(\mathbf{0}, \sigma_q^2 \mathrm{I}_\infty)$ with $\sigma_q^2 := \beta(R)\sigma_e^2$, where σ_e^2 is the steady-state component variance of $\mathbf{e}_n$, and $\beta(R) := \frac{1-\alpha(R)}{\alpha(R)}$. The "quantization noise" vector-process $\mathbf{Q}_n$ is temporally independent. In the above representation of the R-D forward test-channel, the quantization noise process is also independent of the source process $\mathbf{X}_{-\infty}^{+\infty}$. As $R \uparrow \infty$, $\widetilde{\mathbf{e}}_n(i) \rightarrow \mathbf{e}_n(i)$ and as $R \downarrow 0$, $\widetilde{\mathbf{e}}_n(i) \rightarrow 0$ for all components i and where the convergence is in the mean-square sense [3].

3. ANALYSIS OF SMOOTHING PERFORMANCE

3.1. Key insight

Conventional theoretical analyses of DPC systems have focused on scalar quantization (corresponding to unit block-length as opposed to infinite block-length in our model) and on high-rate asymptotics. At high rates, the quantized prediction error is very nearly equal to the unquantized prediction error. This can also be seen in our model by taking R to infinity and noting that $\mathbb{E}|\mathbf{e}_n(i) - \widetilde{\mathbf{e}}_n(i)|^2 \rightarrow 0$ for all components i. Since the *unquantized* causal MMSE prediction errors are temporally independent (by the orthogonality principle [9, 3]), it follows that at high rates, there is little to be gained by non-causal temporal smoothing at the decoder because the future quantized prediction errors have little information overlap (correlation) with the current block. This is the main reason why conventional analyses of DPC systems have largely ignored the role of non-causal smoothing at the decoder and have almost exclusively focused on purely causal decoders that perform causal filtering. On the other hand, when the rate is very low, i.e., $R \downarrow 0$, the quantization noise $\mathbf{Q}_n$ *dominates* any useful information overlap that exists between the current and future blocks. Hence, in this regime too, there is little gain in the R-D performance to be realized by non-causal smoothing at the decoder. This informal analysis suggests that there is possibly a regime of *medium* rates for which the gains from non-causal smoothing can be quite significant. The goal of this work is to quantify the gains and identify the regimes of the source correlation $|\rho|$ and rate R where the gains are substantial.

3.2. Derivation of MSE

Let $m \geq 0$ be the order or length of the non-causal smoothing operation at the decoder which represents the size of the decoder buffer or equivalently the relative processing delay before the "display" of the current block. Since all the underlying processes are jointly Gaussian and the distortion criterion is MSE, the MMSE estimate of $\mathbf{X}_n$ based on $\widetilde{\mathbf{e}}_{-\infty}^{n+m}$ is a *linear* combination of the quantized prediction errors [9] and coincides with the MMSE estimate of $\mathbf{X}_n$ based on $\mathbf{Y}_{-\infty}^{n+m}$ where $\mathbf{Y}_n := \mathbf{X}_n + \mathbf{Q}_n$ for all $n \in \mathbb{Z}$. Let $\mathbf{Y}_n^{(m)} := \mathbf{Y}_{n+m}, \forall n \in \mathbb{Z}$ denote the m-step advanced $\mathbf{Y}$ process. Since the component processes are independent and identically evolving, we can restrict attention to *any* scalar-valued component (for definiteness component-1) for all per-component MSE calculations. Let $X_n, Q_n, e_n, \widetilde{e}_n, Y_n, Y_n^{(m)}$, etc., denote the respective component-1 processes. For all $n \in \mathbb{Z}$ we have

$$\begin{aligned} R_X(n) := \mathbb{E}\left[X_n X_{n+n'}\right] &= \sigma_X^2\, \rho^{|n|}, \\ R_{XY}(n) := \mathbb{E}\left[X_{n+n'} Y_n\right] &= \sigma_X^2\, \rho^{|n|}, \\ R_Y(n) := \mathbb{E}\left[Y_n Y_{n+n'}\right] &= \sigma_X^2\, \rho^{|n|} + \beta\, \sigma_e^2\, \delta(n), \\ R_{Y^{(m)}}(n) &= R_Y(n), \end{aligned}$$

where $\delta(0) = 1$ and $\delta(n) = 0$ otherwise. Hence, the z-transforms of the above auto- and cross-correlation functions are given by [3]

$$\begin{aligned} S_X(z) = S_{XY}(z) &= \frac{\sigma_X^2(1-\rho^2)}{(1-\rho z)\,(1-\rho z^{-1})}, \\ S_Y(z) = S_{Y^{(m)}}(z) &= c^2\,\frac{\left(1-bz^{-1}\right)(1-bz)}{(1-\rho z^{-1})\,(1-\rho z)}, \\ S_{XY^{(m)}} &= z^{-m} S_{XY}(z), \end{aligned}$$

for all $z \in \mathbb{C}$ where,

$$c^2 \quad := \quad \frac{\rho}{b}\,\sigma_e^2\,\beta(R),$$

and b is the solution to

$$b + \frac{1}{b} \quad = \quad \rho + \frac{1}{\rho} + \frac{1}{\rho\beta(R)}\frac{\sigma_X^2}{\sigma_e^2}, \tag{1}$$

satisfying $0 < |b| < |\rho| < 1$ and $\mathrm{sign}(b) = \mathrm{sign}(\rho)$.

If $S(z) = \sum_{k \in \mathbb{Z}} s(k) z^{-k}$, let $[S(z)]^c := \sum_{k \geq 0} s(k) z^{-k}$ and $[S(z)]^{nc} := \sum_{k \leq 0} s(k) z^{-k}$ respectively denote the causal

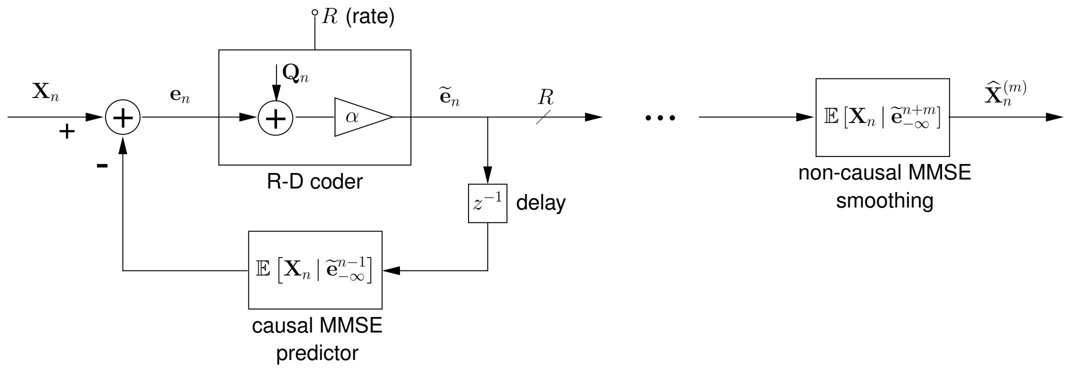

Fig. 1. **Steady-state functional block-diagram of a differential predictive coding system:** *The source $\{\mathbf{X}_n\}_{n=-\infty}^{+\infty}$ is a stationary, AR-1, Gaussian vector-process with independent and identically distributed component processes and temporal correlation ρ. Mean squared error (MSE) is the distortion criterion. The encoder forms the causal MMSE estimate $\mathbb{E}[\mathbf{X}_n|\widetilde{\mathbf{e}}_{-\infty}^{n-1}]$ of the current block $\mathbf{X}_n$ based on all the encoded information available to the decoder till that point. The prediction error for the current block $\mathbf{e}_n$, which is Gaussian, is coded by an ideal R-D coder operating at rate R. The decoder forms a non-causal MMSE estimate $\widehat{\mathbf{X}}_n^{(m)}$ of the current block based on all the information available till time $n+m$ where $m \geq 0$. The statistical relationship between the Gaussian prediction error and its rate-R quantized representation is given by $\widetilde{\mathbf{e}}_n = \alpha(\mathbf{e}_n + \mathbf{Q}_n)$ where $\alpha = 1 - 2^{-2R}$, and $\mathbf{Q}_n \sim$ i.i.d$\mathcal{N}(\mathbf{0}, \sigma_q^2 I_\infty)$ is independent of $\mathbf{X}_{-\infty}^{+\infty}$ with $\sigma_q^2 = \sigma_e^2(\alpha^{-1} - 1)$ where σ_e^2 is the steady-state component variance of $\mathbf{e}_n$. The MSE is a function of σ_X^2, ρ, R, and m. The goal of this work is to develop a quantitative understanding of the gains in MSE, if any, in terms of the smoothing-latency m.*

and non-causal parts of $S(z)$. Let $S(z)$ be the z-transform of an autocorrelation sequence. If $S(z)$ is a rational function of z then it admits the following decomposition: $S(z) = S^+(z)S^+(z^{-1})$ where $S^+(z)$ is causal and causally invertible [3]. Define

$$S_{\text{error}}(z) := S_X(z) - \frac{S_{XY^{(m)}}(z)}{S^+_{Y^{(m)}}(z^{-1})}\left[\frac{S_{XY^{(m)}}(z^{-1})}{S^+_{Y^{(m)}}(z)}\right]^{nc}.$$

The MSE for an m-length non-causal smoothing is given by

$$\text{MSE}\left(\sigma_X^2, \rho, R, m\right) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} S_{\text{error}}\left(e^{j\omega}\right) \mathrm{d}\omega,$$

or equivalently, by the constant coefficient in the Laurent-series expansion of $S_{\text{error}}(z)$ [3]. The expression for MSE can be obtained in closed-form after a fairly tedious series of algebraic steps that are omitted here. The final expression for MSE is given by:

$$\text{MSE}(m) = \text{MSE}(\infty)\left[1 + \frac{(|\rho| - |b|)}{(1 - |\rho|\,|b|)}|b|^{2m+1}\right], \quad (2)$$

$$\text{MSE}(\infty) = \sigma_X^2 \frac{|b|\left(1-\rho^2\right)}{|\rho|\left(1-b^2\right)}. \quad (3)$$

In the steady state it can be shown that

$$\text{MSE}(0) = \mathbb{E}(e_n - \widetilde{e}_n)^2 = \sigma_e^2 \cdot 2^{-2R}. \quad (4)$$

Solving for b from (1), (2), (3), and (4), we obtain

$$b(\rho, R) = \rho 2^{-2R} \quad (5)$$

4. SMOOTHING-LATENCY VERSUS PERFORMANCE TRADEOFFS

4.1. Relative MSE-improvement: potentially how much?

We are now in a position to analyze the improvement in the MSE, if any, due to non-causal smoothing. From (2) observe that the MSE decreases exponentially fast with m from MSE(0) to MSE(∞) with exponent $2\ln(1/|b|)$. Consider the relative improvement in MSE over no smoothing, i.e., $m = 0$, given by

$$\frac{\text{MSE}(0) - \text{MSE}(\infty)}{\text{MSE}(0)} = \frac{|b|\left(|\rho| - |b|\right)}{1 - b^2}. \quad (6)$$

For a given $|\rho| \in (0, 1)$, from (5), (i) As $R \downarrow 0$, $b \to \rho \Rightarrow$ relative MSE-improvement $\to 0$. Hence at very low rates, there is little to be gained by non-causal smoothing. However, the little relative gain that is available can be attained exponentially fast with exponent no poorer than $2\ln\frac{1}{|\rho|}$. (ii) As $R \uparrow \infty$, $b \to 0 \Rightarrow$ relative MSE-improvement $\to 0$. Hence, at very high rates too, there is little to be gained by non-causal smoothing and in addition, the rate of decrease of MSE with m goes to zero. To find the rate at which the relative gain is maximum, set the derivative of the relative MSE-improvement in (6) with respect to R to zero to obtain

$$R_{\max}(|\rho|) = \frac{1}{2}\log_2\left(1 - \sqrt{1-\rho^2}\right)$$

bits per component (pixel). The maximum relative improvement in MSE corresponding to $R_{\max}$ is given by $\frac{1}{2}(1 - \sqrt{1-\rho^2})$ which is in turn maximized at $\rho = 1$ for a maximum relative improvement of $\frac{1}{2}$. Hence,

$$\frac{1}{2}\text{MSE}(0) \leq \text{MSE}(\infty) \leq \text{MSE}(0),$$

i.e., there can be as much as a $10\log_{10}(2) \approx 3$ dB improvement in the peak signal to noise ratio (PSNR) given by $10\log_{10}(\sigma_X^2/\text{MSE})$ in decibels. The percentage relative MSE-reduction for different values of $|\rho|$ and R are plotted in Fig. 2. Note that the relative MSE-improvement does not depend on σ_X^2.

4.2. Relative MSE-improvement: how fast?

We had noted in the last subsection that the MSE reduces exponentially fast with the smoothing latency m with exponent $2\ln(1/|b|)$.

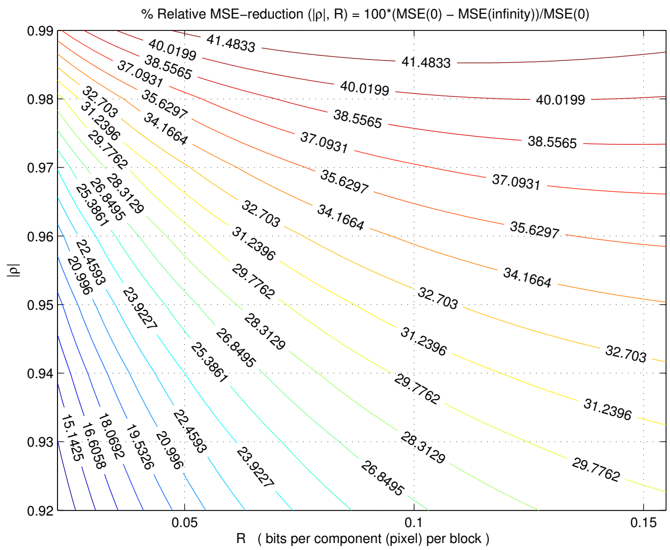

Fig. 2. **Percentage relative MSE-reduction:** *For each* $|\rho| \in (0, 1)$, *the improvement is maximum at the rate* $\frac{1}{2}\log_2(1 + \sqrt{1-\rho^2})$ *and goes to zero for* $R \to 0$ *and for* $R \to \infty$ *(not visible in this figure). The maximum gain is* 50% *for* $|\rho|$ *close to* 1. *For any given rate* R, *the relative gain is greater for a larger value of* $|\rho|$.

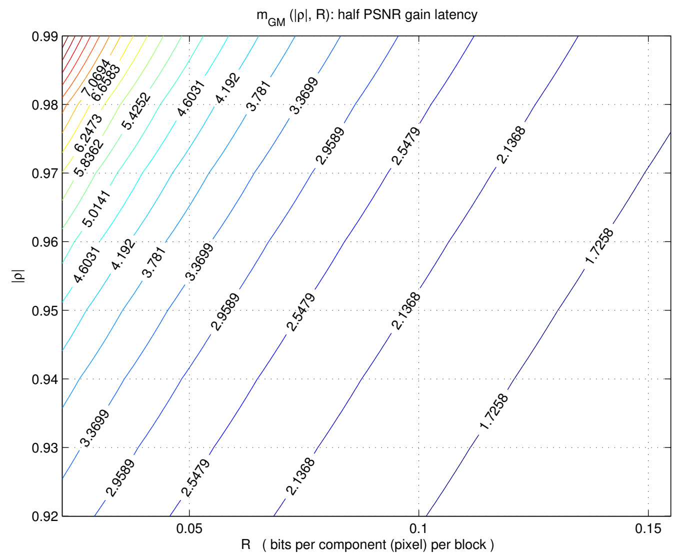

Fig. 3. $m_{GM}(|\rho|, R)$: *From Fig. 2, at rate* $R = 0.1$ *and strong temporal correlation* $|\rho| \approx 0.98$, *there is about a* 40% *relative MSE-reduction corresponding to a potential PSNR improvement of PSNR*(∞) − *PSNR*$(0) \approx 2.2$*dB. Half of this gain, i.e.,* 1.1*dB can be attained with a latency of* $m_{GM} \approx 3$ *as seen above.*

We also have an idea of the extent of the gains that can be achieved at different rates R and correlation strengths $|\rho|$. We would now like to quantify the smoothing latency m needed to realize some of these gains. Let m_{AM} be the latency needed to realize half the gains in MSE, i.e.,

$$\text{MSE}(m_{\text{AM}}) = \frac{1}{2}\left(\text{MSE}(0) + \text{MSE}(\infty)\right).$$

From (2) and (3) it follows that

$$m_{\text{AM}} = \frac{\log 2}{2\log\left(\frac{1}{|b|}\right)}.$$

As an alternative measure of the speed of convergence, let m_{GM} be the latency needed to realize half the gains in PSNR, i.e.,

$$\text{PSNR}(m_{\text{GM}}) = \frac{1}{2}\left(\text{PSNR}(\infty) + \text{PSNR}(0)\right).$$

It can be verified that $\text{MSE}(m_{\text{GM}}) = \sqrt{\text{MSE}(0)\cdot\text{MSE}(\infty)}$. Solving for m_{GM} using (2) and (3) we obtain

$$m_{\text{GM}} = \frac{m_{\text{AM}}}{\log 2}\log\left(1 + \sqrt{\frac{1-|b|^2}{1-|\rho|\,|b|}}\right).$$

Note that $m_{AM} \leq m_{\text{GM}}$ since the arithmetic mean of MSEs is never smaller than the geometric mean. Figure 3 shows the variation of m_{GM} with the rate R and the correlation strength $|\rho|$.

5. CONCLUDING REMARKS

The take-away message of this analysis is that "substantial" performance gains (e.g., 1–1.5 dB) can be realized with "small" smoothing latency (e.g., m = 2, 3) for sources with a "strong temporal correlation" (e.g., ρ = 0.98) coded at low bitrates (e.g., $R = 0.1$ bits per pixel). The gains vanish at very high and very low rates. Extensions to more sophisticated source models and finite block-size are interesting directions for taking this work further.

6. REFERENCES

[1] N. Farvardin and J. W. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources,," *IEEE Trans. on Information Theory*, vol. 31, no. 3, pp. 402–418, May 1985.

[2] O. G. Guleryuz and M. T. Orchard, "On the DPCM Compression of Gaussian Auto-Regressive Sequences," *IEEE Trans. on Information Theory*, vol. 47, no. 3, pp. 945–956, March 2001.

[3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Companies, USA, 1965.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.

[5] C. Stiller and J. Konrad, "Estimating motion in image sequences," *IEEE Sig. Proc. Magazine*, vol. 16, no. 4, pp. 70–91, Jul 1999.

[6] J. R. Ohm, "Three-Dimensional Subband Coding with Motion Compensation," *IEEE Trans. Image Proc.*, vol. 3, pp. 559–571, Sep 1994.

[7] S-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Image Communications*, vol. 16, pp. 705–724, May 2001.

[8] D. Taubman and A. Secker, "Highly scalable video compression with scalable motion coding," in *Proc. IEEE Int. Conf. on Image Proc.*, Barcelona, Spain, Sep 2003, IEEE, vol. 3, pp. 273–276.

[9] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer–Verlag, New York, NY, 1994.

[10] S. S. Pradhan, "Approximation of test channels in source coding," in *Proc. Conference on Information Sciences and Systems*, Princeton University, Mar 2004.