CHANNEL-AWARE RATE-DISTORTION OPTIMIZED LEAKY MOTION PREDICTION

Zhen Li and Edward J. Delp

Video and Image Processing Laboratory(VIPER)
School of Electrical and Computer Engineering
Purdue University,West Lafayette, Indiana, USA

ABSTRACT

Leaky motion prediction techniques have been developed as a way to trade-off between video coding efficiency and drift error resilience. In this paper, we present a statistical analysis of leaky motion prediction in the presence of channel errors. We assume the encoder has some basic knowledge of the channel such as the channel error pattern and error rate. We derive a closed-form expression of the rate distortion function and find the analytic solution for the leaky motion prediction parameter. Two examples are presented to demonstrate the results.

1. INTRODUCTION

In motion prediction hybrid video coding, the encoder and decoder are supposed to have access to the same motion reference information. A drift error occurs when this condition is not satisfied. This can happen when there are errors in the channel during the transmission of the reference information or the channel bandwidth is not adequate to represent all the reference information and has to discard part or all of the data stream. In case of drift, the reconstructed video quality can deteriorate quickly due to error propagation in the motion compensation until the next INTRA frame (I-frame) occurs.

Many techniques have been investigated to limit or stop drift errors. Forward error correction(FEC) techniques are widely used in practice and can prevent the drift effectively[1]. However, FEC-based approaches generally require extra coding complexity, which is critical in real-time or handheld video applications. FEC-based techniques require bandwidth for the redundant information and may significantly reduce coding efficiency. Another approach is to use layered scalable video coding[2][3]. A lower quality reconstruction, which is referred to as the base layer, requiring a lower data rate, is used as the reference. If the decoder receives more bits than the base layer, it uses the extra bits for higher quality reconstruction. The higher quality reconstruction will not be used as the reference. Therefore, the encoder and decoder only need to guarantee the quality of the base layer to prevent drift errors. A lower quality reference will lead to larger residual image entropy resulting in a higher data rate and/or decreased coding efficiency.

Leaky prediction based techniques[4] [5] have been proposed to trade-off between coding efficiency and drift error resilience. Leaky prediction uses a fraction, referred to as the leaky factor α, of the difference between the higher quality reconstruction and the lower quality reconstruction along with the lower quality reconstruction as a reference. It is obvious that the selection of the leaky factor will greatly affect the performance of the motion prediction. When $\alpha = 0$ it is essentially the conventional SNR layered coding, where the greatest error resilience is achieved; while when $\alpha = 1$ it is the single layer coding, where coding efficiency is maximized.

The selection of an optimal leaky factor is a difficult task due to the lack of a well defined channel model and frame dependence. A motion prediction rate distortion analysis was proposed in [6] and extended in [7]. This approach is continued in [8] [9] to address the leaky parameter problem in scalable video coding. Our work in this paper differs from the previous work in the following three aspects. First, we model the video signal as a first-order Markov-like vector sequences. Based on this model, we derive a recursive expression of the distortion. Second, we assume that the encoder is channel-aware, i.e., it has the knowledge of the channel conditions such as the error pattern and the error rate. We shall consider the channel errors explicitly in the rate distortion function. Finally, since the average data rate in many video coding applications is generally well under 1 bit per pixel, we extend our work in [10] and exploit the high resolution quantization analysis in [11]. We derive a analytic solution for the leaky parameter in this paper. It should be noted that our results essentially extend the 1-D analysis in [12] to 2-D video signals.

2. STATISTICAL ANALYSIS OF MOTION PREDICTION WITH CHANNEL ERRORS

In hybrid video coding, the n-th input frame of size $M \times N$ is divided into blocks of size $L \times L$. An orthogonal $L \times L$ transform T is then obtained for each block. The motion prediction and motion compensation operation on the n-th frame is a 2-D nonlinear filter denoted as P_n.

We denote $x(i, j, n)$ and $y(i, j, n)$ as the n-th input-frame at the encoder and reconstructed frame at the decoder respectively, where (i, j) is the spatial coordinates of the pixel. The difference image is

$$\varepsilon(i, j, n) = y(i, j, n) - x(i, j, n) \tag{1}$$

Denote the inner product of vectors $a(n)$ and $b(n)$ by

$$\Phi[a(n), b(n)] = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} a(i, j, n) \cdot b(i, j, n) \tag{2}$$

The mean square error(MSE) is then

$$D(n) = \Phi[\varepsilon(n), \varepsilon(n)] \tag{3}$$

Our goal is to minimize the MSE given the channel conditions. We assume that the video sequence can be approximated by a first-order Markov-like model

This work was supported by the Indiana 21^{st} Century Research and Technology Fund. Address all correspondence to Prof. E. J. Delp, ace@ecn.purdue.edu.

$$x(n) = \rho\tilde{x}(n-1) + w(n) \quad (4)$$

where $\tilde{x}(n-1)$ is the motion compensated version of the $(n-1)$-th reconstructed frame $x_q(n-1)$ at the encoder, i.e.,

$$\tilde{x}(n-1) = P_n(x_q(n-1)) \quad (5)$$

We focus our work on the selection of the leaky factor,α, and do not consider the base layer data rate selection. Without loss of generality, we assume the base layer rate is zero. Hence the reference image and the residual image are

$$\hat{x}(n) = \alpha\tilde{x}(n-1) \quad (6)$$

$$e(n) = x(n) - \hat{x}(n) \quad (7)$$

Let the quantization operation be denoted as $Q[\cdot]$ and the inverse quantization operation as $Q^{-1}[\cdot]$. The quantized residual image is then

$$q_{enc}(n) = Q[e(n)] \quad (8)$$

The dequantized residual image at the encoder is

$$e_{enc}(n) = Q^{-1}[q_{enc}(n)] = e(n) + d_q(n) \quad (9)$$

where $d_q(n)$ is referred to as the quantization error. Hence the reconstructed frame at the encoder becomes

$$x_q(n) = \hat{x}(n) + e_{enc}(n) = x(n) + d_q(n) \quad (10)$$

Let $q_c(n)$ be the effect of the channel errors and assume the error effects are additive, then the quantized residual image received at the decoder can be represented as

$$q_{dec}(n) = q_{enc}(n) + q_c(n) \quad (11)$$

The dequantized residual image at the decoder is

$$\begin{aligned} e_{dec}(n) &= Q^{-1}[q_{dec}(n)] = Q^{-1}[q_{enc}(n) + q_c(n)] \\ &= e(n) + d_q(n) + d_c(n) \end{aligned} \quad (12)$$

Hence the reconstructed n-th frame is

$$y(n) = \alpha P_n[y(n-1)] + e_{dec}(n) \quad (13)$$

And the difference image is

$$\begin{aligned} & \varepsilon(n) = y(n) - x(n) \\ = \;& \alpha P_n[y(n-1)] + e_{dec}(n) - x(n) \\ = \;& \alpha P_n[y(n-1)] + e(n) + d_q(n) + d_c(n) - x(n) \\ = \;& \alpha P_n[y(n-1) - x_q(n-1)] + e(n) + d_q(n) \\ + \;& d_c(n) + \alpha P_n[x_q(n-1)] - x(n) \\ = \;& \alpha P_n[y(n-1) - x_q(n-1)] + d_q(n) + d_c(n) \\ = \;& \alpha P_n[y(n-1) - x(n-1) - d_q(n-1)] + d_q(n) + d_c(n) \\ = \;& \alpha P_n[\varepsilon(n-1)] - \alpha P_n[d_q(n-1)] + d_q(n) + d_c(n) \end{aligned} \quad (14)$$

Continuing this procedure recursively, we have

$$\begin{aligned} \varepsilon(n) &= \alpha^n \cup_{i=1}^{n} P_i[\varepsilon(0) - d_q(0) - d_c(0)] \\ &+ \sum_{k=0}^{n-1} \alpha^k \cup_{i=0}^{k} P_{n-i}[d_c(n-k)] + d_q(n) \end{aligned} \quad (15)$$

where $\cup_{i=m}^{n} P_i = P_m[P_{m+1}[...[P_n[\cdot]]]]$.

Since the 0-th frame is an INTRA frame, its difference image is not dependent on any previous frames, i.e., $\varepsilon(0) = d_q(0) + d_c(0)$. Hence, we can decouple the difference image as

$$\varepsilon(n) = \varepsilon_c(n) + \varepsilon_q(n) \quad (16)$$

where $\varepsilon_c(n)$ is the distortion due to channel erros and $\varepsilon_q(n)$ is distortion due to quantization errors, i.e.,

$$\varepsilon_c(n) = \sum_{k=0}^{n-1} \alpha^k \cup_{i=0}^{k} P_{n-i}[d_c(n-k)] \quad (17)$$

$$\varepsilon_q(n) = d_q(n) \quad (18)$$

Suppose the channel error and quantization error are uncorrelated,

$$\Phi[\varepsilon(n), \varepsilon(n)] = \Phi[\varepsilon_q(n), \varepsilon_q(n)] + \Phi[\varepsilon_c(n), \varepsilon_c(n)] \quad (19)$$

We note that the motion compensation operation P_n is a relocation of the pixels inside a frame, we can assume P_n does not change the inner product of two images, i.e.,

$$\Phi[d_c(n), d_c(n)] = \Phi[P_n[d_c(n)], P_n[d_c(n)]] \quad (20)$$

Hence

$$\begin{aligned} & \Phi[\varepsilon_c(n), \varepsilon_c(n)] \\ = \;& \Phi[\sum_{k=1}^{n-1} \alpha^k d_c(n-k), \sum_{k=1}^{n-1} \alpha^k d_c(n-k)] \\ = \;& \Phi[\sum_{k=1}^{n-1} \alpha^k (e_{dec}(n-k) - e_{enc}(n-k)), \\ & \sum_{k=1}^{n-1} \alpha^k (e_{dec}(n-k) - e_{enc}(n-k))] \\ = \;& \sum_{k=0}^{n-1} \alpha^{2k} \Phi[e_{dec}(n-k), e_{dec}(n-k)] \\ & + \sum_{k=0}^{n-1} \alpha^{2k} \Phi[e_{enc}(n-k), e_{enc}(n-k)] \\ & - 2\sum_{k=0}^{n-1} \alpha^{2k} \Phi[e_{dec}(n-k), e_{enc}(n-k)] \\ = \;& (\Phi[e_{dec}(n), e_{dec}(n)] + \Phi[e_{enc}(n), e_{enc}(n)] \\ & - 2\Phi[e_{enc}(n), e_{dec}(n)]) \cdot \\ & (\frac{1-\alpha^{2n}}{1-\alpha^2}(1 - \delta(\alpha - 1)) + n\delta(\alpha - 1)) \end{aligned} \quad (21)$$

where $\delta(\alpha) = 1$ if $\alpha = 0$, otherwise $\delta(\alpha) = 0$. The third equality of (21) comes from the observation that the $i-th$ and $j-th$ residual frames are uncorrelated if $i \neq j$; and the fourth equality follows the assumption that the residual frames in a video sequence are stationary, i.e., $\Phi[e(n-k), e(n-k)] = \Phi[e(n), e(n)], \forall k$.

On the other hand, the quantization distortion can be represented by [13]

$$\Phi[\varepsilon_q(n), \varepsilon_q(n)] = \Phi[d_q(n), d_q(n)] = \xi^2 \sigma_e^2(n) 2^{-2R} \quad (22)$$

where ξ^2 is a constant depending on the quantization mechanism and also the video sequence characteristics. And

$$\begin{aligned} & \sigma_e^2(n) = \Phi[x(n) - \alpha\tilde{x}(n-1), x(n) - \alpha\tilde{x}(n-1)] \\ = \;& \Phi[(\rho - \alpha)\tilde{x}(n-1) + w(n), (\rho - \alpha)\tilde{x}(n-1) + w(n)] \\ = \;& [(\rho - \alpha)^2 + (1 - \rho^2)]\Phi[\tilde{x}(n), \tilde{x}(n)] \end{aligned} \quad (23)$$

where $\Phi[w(n), w(n)] = (1 - \rho^2)\Phi[\tilde{x}(n), \tilde{x}(n)]$ follows from (4) and the stationarity assumption of the sequence. With (19) to (23),

$$\begin{aligned} D(n) &= \xi^2[(\rho - \alpha)^2 + (1 - \rho^2)]\Phi[\tilde{x}(n), \tilde{x}(n)]2^{-2R} + \\ & (\Phi[e_{dec}(n), e_{dec}(n)] + \Phi[e_{enc}(n), e_{enc}(n)] - \\ & 2\Phi[e_{enc}(n), e_{dec}(n)])(\frac{1-\alpha^{2n}}{1-\alpha^2}(1 - \delta(\alpha - 1)) + n\delta(\alpha - 1)) \end{aligned} \quad (24)$$

this is the closed-form expression of the rate-distortion function for motion prediction with drift.

Up to this point, the analysis is done in the pixel domain, where the quantization noise and channel errors are assumed to be added directly to each pixel. However, in most video coding frameworks the residual image is first transformed before quantization and channel transmission. Our analysis may raise a concern of extending our results into the transform domain. Since most transforms used in practical coding frameworks can be approximated with orthogonal transforms, which are variance preserving, we can use the analysis developed in the pixel domain for the transform domain.

3. EXAMPLES

3.1. Error Free Channel

We first look at the simplest case where there is no channel error,

$$e_{dec}(n) = e_{enc}(n) \tag{25}$$

$$D(n) = \xi^2[(\rho-\alpha)^2 + (1-\rho^2)]\Phi[\tilde{x}(n), \tilde{x}(n)]2^{-2R} \tag{26}$$

Use Lagrangian optimization on (26)

$$\frac{\partial D(n)}{\partial \alpha} = 0 \Longrightarrow \alpha_{opt} = \rho \tag{27}$$

which is consistent with linear prediction theory.

3.2. Random Bit Error P_b

We now consider another case where the channel is no longer free of error. Instead, it has random bit errors with bit error rate(BER) of P_b, which is a common model for wireless channels.

Let $E_{enc}(n) = T[e_{enc}(n)]$ and $E_{dec}(n) = T[e_{dec}(n)]$ be the transform coefficients. Assume that each transform coefficient $E_{enc}(i,j,n)$ in $E_{enc}(n)$ and $E_{dec}(i,j,n)$ in $E_{dec}(n)$ are represented with $B(i,j,n)$ bits respectively, and

$$\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} B(i,j,n) = R \tag{28}$$

where R is the average bit budget. Then

$$E_{enc}(i,j,n) = V(i,j,n)\sum_{k=1}^{B(i,j,n)} E_{enc}(i,j,n,k)2^{-k} \tag{29}$$

where $V(i,j,n)$ represents the range of transform coefficients at the same frequency and $E_{enc}(i,j,n,k)$ is the k-th bit in the binary representation of $E_{enc}(i,j,n)$. Similarly,

$$E_{dec}(i,j,n) = V(i,j,n)\sum_{k=1}^{B(i,j,n)} E_{dec}(i,j,n,k)2^{-k} \tag{30}$$

We now assume that the probability that $E(i,j,n,k)$ takes positive or negative ones are equal. Hence the autocorrelation for $E(i,j,n)$ is

$$\Phi[E_{enc}(i,j,n), E_{enc}(i,j,n)] = \frac{1-2^{-2B(i,j,n)}}{3}V^2(i,j,n) \tag{31}$$

Since the bit error is random,

$$\begin{aligned}&\Phi[E_{dec}(i,j,n), E_{dec}(i,j,n)] \\ =\ & \frac{1-2^{-2B(i,j,n)}}{3}(1-P_b)V^2(i,j,n)\end{aligned} \tag{32}$$

$$\begin{aligned}&\Phi[E_{enc}(i,j,n), E_{enc}(i,j,n)] - \Phi[E_{dec}(i,j,n), E_{dec}(i,j,n)] \\ =\ & \frac{1-2^{-2B(i,j,n)}}{3}P_bV^2(i,j,n)\end{aligned} \tag{33}$$

Although more sophisticated models, such as a Laplacian model, can be used for the transform coefficients, we assume that the transform coefficients are uniformly distributed, then

$$\sigma^2_{E_{enc}(i,j,n)} = \frac{V^2(i,j,n)}{3} \tag{34}$$

With (34)

$$\begin{aligned}&\Phi[E_{enc}(i,j,n), E_{enc}(i,j,n)] - \Phi[E_{dec}(i,j,n), E_{enc}(i,j,n)] \\ =\ & P_b(1-2^{-2B(i,j,n)})\sigma^2_{E_{enc}(i,j,n)}\end{aligned} \tag{35}$$

We then sum over all transform coefficients

$$\begin{aligned}&\Phi[E_{enc}(n), E_{enc}(n)] - \Phi[E_{dec}(n), E_{enc}(n)] \\ =\ & \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} P_b(1-2^{-2B(i,j,n)})\sigma^2_{E_{enc}(i,j,n)}\end{aligned} \tag{36}$$

A classical method for bit allocation of quantized coefficient under an average rate constraint R is given in [11] using the high resolution quantization approximations. The optimal bit allocation is

$$B(i,j,n) = R + \frac{1}{2}\log_2\frac{\sigma^2_{E_{enc}(i,j,n)}}{\rho^2} \tag{37}$$

where

$$\rho^2 = (\prod_{i=0}^{M-1}\prod_{j=1}^{N-1}\sigma^2_{E_{enc}(i,j,n)})^{\frac{1}{MN}} \tag{38}$$

is the geometric mean of the variance of the random variables. Substitute (37) into (36) and use the variance preserving property,

$$\sigma_e^2 = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\sigma^2_{e_{enc}(i,j,n)} = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\sigma^2_{E_{enc}(i,j,n)} \tag{39}$$

we have

$$\begin{aligned}&\Phi[E_{enc}(n), E_{enc}(n)] - \Phi[E_{dec}(n), E_{dec}(n)] \\ =\ & P_b\sigma_e^2[1-2^{-2R}r]\end{aligned} \tag{40}$$

where r is the ratio of geometric mean to the arithmetic mean for the residual image variance, which is defined as

$$r = \frac{(\prod_{i=0}^{M-1}\prod_{j=0}^{N-1}\sigma^2_{E_{enc}}(i,j,n))^{\frac{1}{MN}}}{\frac{1}{MN}(\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\sigma^2_{E_{enc}}(i,j,n))} \tag{41}$$

It is well-known that this ratio is always equal to or less than 1. Hence with (24) and (40)

$$\begin{aligned}D(n) \quad =\quad & \Phi[\tilde{x}(n), \tilde{x}(n)](\frac{1-\alpha^{2n}}{1-\alpha^2}(\alpha^2 - 2\alpha\rho + 1)P_b \\ & (1-2^{-2R}r) + \xi^2 2^{-2R}((\rho-\alpha)^2 + (1-\rho^2)))\end{aligned} \tag{42}$$

Use Lagrangian optimization on (42)

$$\frac{\partial D(n)}{\partial \alpha} = 0 \tag{43}$$

We now show several simulation results based on this analysis. For the sake of simplicity, $\xi = 1$ and $r = 0.90$. We notice in our simulations that although r should actually be calculated from the residual frames, the selection of r does not affect the results very much because it is weighted by 2^{-2R}.

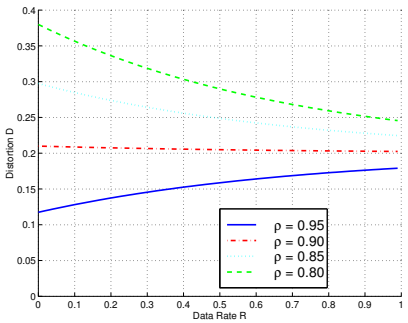

Fig. 1. Distortion rate function when $P_b = 0.20$.

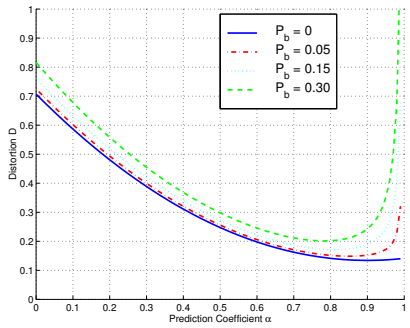

Fig. 2. Distortion D vs. prediction coefficient α when $R = 0.25$ bit per pixel, $\rho = 0.90$ and P_b varies.

The $D(R)$ function is shown in Fig. 1. Here we use the correlation coefficient ρ as the leaky prediction parameter, $\rho = \alpha$. As we can see, in presence of channel errors, the increase in data rate does not necessarily reduce the distortion, in particular, for video sequences with high inter frame correlation. Fig. 2 shows the results of distortion with different leaky prediction parameters. The optimal α that minimizes the distortion deviates from the correlation coefficient ρ with the increase in channel errors. The optimal leaky prediction coefficients corresponding to $P_b = 0, 0.05, 0.15, 0.30$ are $\alpha = 0.90, 0.85, 0.81, 0.78$ respectively and were obtained from (42) The relationship between α and P_b is given in Fig. 3. The curves in Fig. 3 show that α deviates from the correlation coefficient in a similar way with the increase of channel error.

4. CONCLUSIONS

In this paper, we presented a statistical analysis of leaky motion prediction in the presence of channel errors. Another important problem in leaky motion prediction is the selection of the base layer data rate. Currently we are investigating a generalized analysis to include the base layer rate in the rate distortion function. We are also examining other channel error patterns, in particular bit-plane coding with packet loss channels.

5. REFERENCES

[1] Y. Wang and Q. Zhu, "Error control and concealment for video communcations: A review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May, 1998.

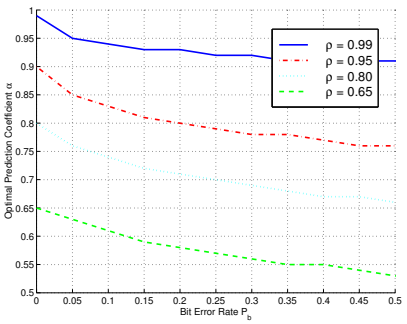

Fig. 3. Optimal prediction coefficient α vs. bit error rate P_b.

[2] ITU-T and ISO/IEC JTC1, "Generic coding of moving pictures and associated audio information - part 2: Video," *ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2)*, November 1994.

[3] ITU-T and ISO/IEC, "Text of committee draft of joint video specification," *ITU-T Recommendation H.264—ISO/IEC 14496-10 AVC(MPEG-4 Part 10)*, 2003.

[4] H. Huang, C. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 372–385, June 2002.

[5] S. Han and B. Girod, "Robust and efficient scalable video coding with leaky prediction," in *Proceedings of the IEEE International Conference on Image Processing*, Rochester, NY, September 22-25, 2002, vol. 2, pp. 41–44.

[6] B. Girod, "The efficiency of motion-compensation prediction for hybrid coding of video sequences," vol. SAC-5, pp. 1140–1154, Aug. 1987.

[7] G. W. Cook, *A Study of Scalability in Video Compression: Rate-Distortion Analysis and Parallel Implementation*, Ph.D. thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, Dec. 2002.

[8] Y. Liu, J. Prades-Nebot, P. Salama, and E. J. Delp, "Rate distortion analysis of leaky prediction layered video coding using quantization noise modelling," in *Proceedings of the IEEE International Conference on Image Processing*, Singapore, Oct.24–27, 2004.

[9] J. Prades-Nebot, G. W. Cook, and E. J. Delp, "Analysis of the efficiency of SNR-scalable strategies for motion compensated video coders," in *Proceedings of the IEEE International Conference on Image Processing*, Singapore, Oct.24–27, 2004.

[10] Z. Li and E. J. Delp, "Statistical motion prediction with drift," in *Proceedings of the SPIE International Conference on Video Communications and Image Processing*, San Jose, CA, Jan.18–22, 2004, vol. 5308, pp. 416–427.

[11] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Norwell, MA, 1991.

[12] J. Essman, *Theory and Applications of DPCM and Comparison of DPCM with Other Time Predictive Techniques*, Ph.D. thesis, Purdue University, West Lafayette, IN, 1972.

[13] N. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, NJ, 1984.