THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 5, May–June 1978
Printed in U.S.A.

On Predictive Quantizing Schemes

By P. NOLL

(Manuscript received October 26, 1977)

*This paper analyzes the performance of various predictive quantizing schemes both for noiseless and noisy channels. The fidelity criterion used to define optimum performance is that of minimum mean-squared error. The first part of this paper compares differential pulse code modulation (DPCM) with a system that lacks the feedback around the quantizer. Such a system (that is called D*PCM in this paper) is actually a pulse code modulation (PCM) system with a prefilter and a postfilter. In the second part of this paper a noise-feedback coding structure is used as a framework for a unified analysis of predictive quantizing schemes with a frequency-weighted mean-squared error as the performance criterion. The last part of this paper extends the analysis to include the effects of channel transmission errors on the overall performance of these predictive quantizing schemes. It is shown that DPCM and D*PCM when appropriately optimized are less sensitive to channel errors than PCM, and that the performances of DPCM and D*PCM are almost identical in the case of high bit-error rates.*

I. INTRODUCTION

Predictive quantizing schemes employ prediction to exploit the inherent redundancy of input signals; the difference between the actual sample of an input signal and its estimate is quantized and transmitted to the receiver in a digital format. If the samples of the input are highly correlated, then the variance of the samples of the difference signal will be significantly less than the variance of the input samples. Hence, the overall error between input and output of the communication system will be lower than that of a conventional pulse code modulation (PCM) system. The first part of this paper compares PCM and two differential (predictive) pulse code modulation systems (see Fig. 1). A noise-feedback coding structure is then used as a framework of a unified analysis of predictive quantizing schemes (including those of Fig. 1) on the basis of a frequency-weighted error criterion. The last part of this paper ex-

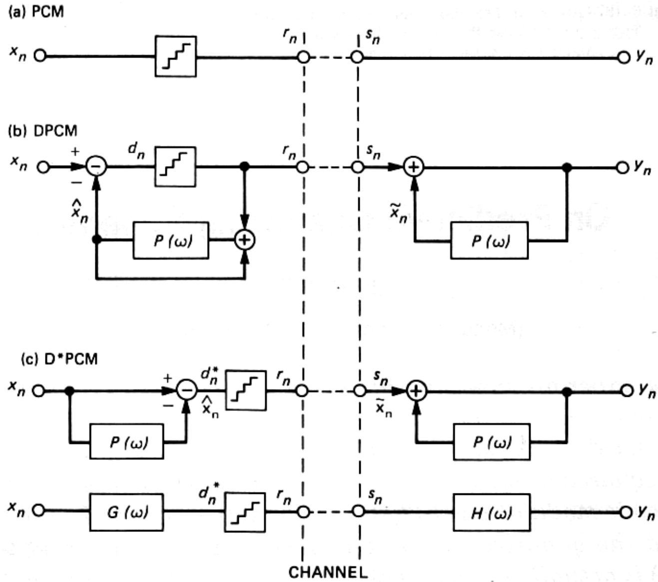

Fig. 1—Structures of PCM, DPCM, and D*PCM coders.

tends the analysis to include the effects of channel transmission errors on the overall performance of these predictive quantizing schemes.

The differential pulse code modulation (DPCM) system has a feedback around the quantizer resulting in a prediction that is based on previous *reconstructed* samples (Fig. 1b). The idea is to base the prediction on information that is also available at the receiver.[1] The D*PCM system is an open-loop quantization scheme,[2] i.e., previous samples of the input are used for predicting the actual input samples (Fig. 1c). The term D*PCM indicates that we have denoted by $d*_n$ the sequence of difference samples in the open-loop system. It is important to realize that coder and decoder calculate different prediction values. We also see, from Fig. 1c, that in this latter scheme the prediction network can be replaced with a more general linear network. The resulting scheme is then actually a PCM scheme with pre- and postfiltering and we shall use this latter scheme to derive bounds for the D*PCM performance.

Let us briefly discuss the differences between DPCM and D*PCM. We denote by $q_n = d_n - r_n$ and $q_n = d*_n - r_n$ the quantization errors of DPCM and D*PCM, respectively, and by r_n the quantized versions of the difference samples.

It is easy to see that the total coding error

$$t_n = x_n - y_n \tag{1}$$

between encoder input x_n and decoder output y_n is given by

$$t_n = q_n \text{ for DPCM} \tag{2}$$

provided the channel is error-free and the decoder uses the same predictor of frequency response $P(\omega)$. On the other hand,

$$t_n = q_n + (\hat{x}_n - \tilde{x}_n) \text{ for D*PCM}, \tag{3}$$

i.e., the total error is increased by the difference between the prediction values $\hat{x}_n$ and $\tilde{x}_n$ of the predictors of encoder and decoder, respectively. Alternatively, we may write

$$t_n = q_n * h_n \text{ for D*PCM} \tag{4}$$

provided that the networks of encoder and decoder are reciprocal. In eq. (4), $*$ denotes discrete-time convolution, and h_n; $n = 0,1,2\ldots$ is the impulse response of the linear decoder network. The result can be easily derived by recognizing that

$$\begin{aligned} t_n &= x_n - y_n \\ &= x_n - r_n * h_n \\ &= x_n - (d^*_n - q_n) * h_n \\ &= q_n * h_n + x_n - x_n * g_n * h_n \\ &= q_n * h_n \end{aligned}$$

since $g_n * h_n = \delta_n$ and $x_n * \delta_n = x_n$ (g_n is the impulse response of the encoder network and δ_n is the Kronecker delta).

Quantization errors are approximately white noise samples if the number of quantizer levels is sufficiently high. Hence in D*PCM the total error is nonwhite noise with each quantization error causing an infinite output sequence. This error propagation effect is also sometimes called error accumulation; this latter term, however, should be used cautiously, because it may seem to imply that D*PCM cannot give any improvement in signal-to-noise ratio (SNR) over PCM, or, even more strongly, that the overall performance can only be degraded.

One purpose of this paper is to analyze and explain the differences between DPCM and D*PCM. The D*PCM coding system has been analyzed by Bodycomb and Haddad[3]; their approach has been based on a predictor optimized for a minimum prediction error variance and they have shown that this predictor as a prefilter for a Gaussian process produces the same total error variance as a PCM system. We shall use the term MMSE predictor to describe such a predictor (that has been optimized for a minimum prediction error variance). We shall see very shortly that this D*PCM scheme can perform better than PCM provided that the predictor is reoptimized; the MMSE predictor is not optimal in this context.

We have already mentioned that D*PCM can be viewed as a PCM scheme with pre- and postfilters. If we model the quantizer in the PCM scheme as an additive noise source, the optimization of this scheme is almost identical to a joint optimization of pre- and postfilters in communication systems in which channel errors are present. Many contributions have been made to this problem both for continuous-time and pulse-modulated communication systems.[4–13] One common result that can be extracted from these papers is that *half-whitening* of the input spectrum minimizes the overall mean-squared error (MSE) between input and output of a communication system with pre- and postfilters in which additive white noise (either caused by a quantization or by channel errors or by both effects) is present. Half-whitening is obtained if the magnitude-squared frequency response is inversely proportional to the square-root of the power density spectrum of the input signal.

PCM schemes with pre- and postfilters do not take into account that in systems with quantizers not only can the input spectrum be shaped to improve the overall performance but that, additionally, the quantization noise spectrum can be shaped as well to further improve the performance of the system. This problem has been discussed in detail by Kimme and Kuo[14] and later by Brainard and Candy[15] on the basis of different coder configurations. These coders have in common a feedback of filtered quantization noise to the input of the quantizer and we shall use the term *noise-feedback coding* (NFC) for this approach. Our analysis differs from earlier contributions in that a power constraint on the quantizer input is not needed. We offer a simplified solution based on well-known results in prediction theory.

It is the aim of this paper to discuss the differences between DPCM and D*PCM both for noise-free and noisy channels and to show how they relate to noise-feedback coding if the basis of the comparison is a frequency-weighted minimum mean-squared error. The organization of this paper is as follows: in Section II we calculate the differences in performance between predictive quantizing schemes with and without feedback around the quantizer. We show that D*PCM can perform better than PCM for all nonwhite input spectra, but that its performance is always below that obtainable with a DPCM scheme. A bound will be derived for the differences in performance between these two schemes, and a first-order Markov source will be used as an example to explain these differences. Section III analyzes noise-feedback coding; its structure is given by a prefilter followed by a quantizer with feedback around the quantizer.[14] We note that PCM, PCM with noise feedback, DPCM, and D*PCM are special cases of this configuration; thus a unified approach is possible. We optimize this coder with a frequency-weighted mean-squared error as the performance criterion and show that the prefilter has to be a whitening filter for the input signal irrespective of the chosen frequency-weighting and that the feedback filter is the MMSE predictor of the weighted input spectrum. We also give frequency-weightings for which the noise-feedback scheme degenerates to DPCM or to D*PCM. Section IV extends the analysis to include the effects of channel transmission errors on the overall performance of predictive quantizing schemes. It is shown that the effects of these errors on total MSE can be significantly smaller than those in PCM. In D*PCM the optimum filters minimize simultaneously quantization and channel error variances and do not depend on the bit-error rate. In the case of DPCM a compromise is needed in order to minimize the total effect of both noise sources on the total MSE. At high bit-error rates the performances of D*PCM and DPCM are almost identical if their prediction networks are identical.

The analyses are made under certain restrictions; first of all, we use the MSE as a performance measure (in Section III a frequency-weighted MSE criterion will be taken into account). Second, the analyses are based on the assumption that the quantizer can be modeled as an additive white noise source. It is known that this model is accurate for quantizers with a sufficiently high number of levels. The model is a poor approximation, however, in the case of coarse quantization, especially if the quantizer input samples are highly correlated. The results of our analysis could be extended to these cases by modifying the model but we do not consider such an extension in order not to obscure the main results. We also assume that the variance of the quantization noise is much smaller than that of the signal to be quantized. Comparison with simulation results will show the range where our rather restrictive assumptions hold.

II. ANALYSIS OF PCM, DPCM, D*PCM

The principal aim in this section is to calculate and to compare the variance of the total errors in PCM, DPCM, and D*PCM. We assume that the input is a sample of a zero-mean stationary random sequence $\{x_n\}$ with autocorrelation function

$$R_x(k) = E[x_n x_{n+k}], \tag{5}$$

power density spectrum (pds)

$$S_x(\omega) = \sum_{k=-\infty}^{\infty} R_x(k) e^{-jk\omega}, \tag{6}$$

and variance

$$\sigma_x^2 = R_x(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) d\omega. \tag{7}$$

$S_x(\omega)$ will be assumed to be a rational function of ω. This is not really restrictive since most spectra of practical interest can be approximated by a rational function. The assumption implies that the pds can be represented by $S_x(\omega) = \eta_x^2 A(\omega)A^*(\omega)$ where $\eta_x^2 > 0$ is a scale factor and $A(\omega)$ is the ratio of two polynomials whose zeros are inside the unit circle of the z-domain (factorization theorem for rational spectra). All power density spectra are nonnegative continuous functions defined for $\omega = [-\pi,\pi]$. Since all signals are represented by stationary random sequences, all filters are necessarily of discrete-time type. The action of the quantizer is represented spectrally as white noise added to the quantizer input signal:

$$S_q(\omega) = \sigma_q^2 = R_q(0). \tag{8}$$

We are interested in the variance

$$\sigma_t^2 = E[t_n^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_t(\omega)d\omega \tag{9}$$

of the total error $t_n = x_n - y_n$ [eq. (1)], where $S_t(\omega)$ is the power density spectrum of the zero-mean random sequence $\{t_n\}$. A frequency-weighting of the total error can easily be added if necessary (see Section III).

2.1 PCM

The total error is identical to the quantization error. Hence we have

$$\sigma_t^2 = \sigma_q^2 = \epsilon_q^2 \cdot \sigma_x^2. \tag{10}$$

The quantizer performance factor ϵ_q^2 depends on the properties of the quantizer and on the probability density function of the signal being quantized; its value is the noise variance generated by a quantization of a unit-variance signal. Table IV in Section IV lists various values for the cases of 1-bit and 2-bit quantizers. For a given quantizer we thus find that

$$\min_{\text{PCM}} \{\sigma_t^2\} = \epsilon_q^2 \cdot \sigma_x^2. \tag{11}$$

2.2 DPCM

The total error is again identical to the quantization error but its variance is now proportional to that of the difference signal:

$$\sigma_t^2 = \sigma_q^2 = \epsilon_q^2 \cdot \sigma_d^2. \tag{12}$$

Remark. Some caution is needed if two coding schemes A and B, e.g., PCM and DPCM, are being compared. We have to take into account differences in the quantizer performance factors ϵ_{qA}^2 and ϵ_{qB}^2 of the two schemes, because the probability density functions of the signals at the corresponding quantizer inputs may differ. For example, let A and B denote PCM and DPCM, respectively. The gain of DPCM over PCM in signal-to-noise ratio is then given as

$$\frac{\sigma_{tA}^2}{\sigma_{tB}^2} = \frac{\epsilon_{qA}^2}{\epsilon_{qB}^2} \cdot \frac{\sigma_x^2}{\sigma_d^2}. \tag{13}$$

The ratio $\epsilon_{qA}^2/\epsilon_{qB}^2$ is very close to unity if the quantizers have a logarithmic characteristic because their performance is relatively independent of signal statistics. The ratio is also close to unity if the samples of the coder input sequences are Gaussian distributed since all coders discussed in this paper employ linear networks which do not affect the Gaussian distribution.

In DPCM systems the prediction is affected by the quantization due to the error feedback, i.e., the predictor uses previous reconstruction values $y_j = x_j - t_j = x_j - q_j; j = n - 1, n - 2, \ldots$ instead of the corresponding input samples x_j. The influence of this feedback on the prediction error variance has been analyzed elsewhere[16–18]; it will be briefly discussed in the following example but will then be neglected in the further analyses. It is known that this simplification is valid if quantizers with at least eight levels are employed. If necessary, we shall use the terms "real" DPCM and "ideal" DPCM for analyses based on prediction with and without error feedback, respectively.

Example 1: We show the influence of the feedback on the prediction error in a DPCM scheme with a first-order predictor of value a. Let $\rho = R_x(1)/\sigma_x^2$ be the normalized mutual correlation between adjacent samples. Without feedback the prediction error variance is

$$\sigma_d^2 = (1 + a^2 - 2a\rho) \cdot \sigma_x^2 \tag{14}$$

with a minimum

$$\min\{\sigma_d^2\} = (1 - \rho^2) \cdot \sigma_x^2 \tag{15}$$

for $a = a_{opt} = \rho$. The DPCM difference signal is $d_n = x_n - ax_{n-1} + aq_{n-1}$. Its variance is given as

$$\sigma_d^2 = \frac{1 - 2a\rho + a^2}{1 - \epsilon_q^2 a^2} \cdot \sigma_x^2 \tag{16}$$

on the assumption of a vanishing correlation between input and quantization error,[16] and its minimum variance is obtained with[19]:

$$a_{opt} = \frac{1 + \epsilon_q^{-2}}{2\rho}\left[1 - \sqrt{1 - \epsilon_q^{-2}\left(\frac{2\rho}{1 + \epsilon_q^{-2}}\right)^2}\,\right]. \tag{17}$$

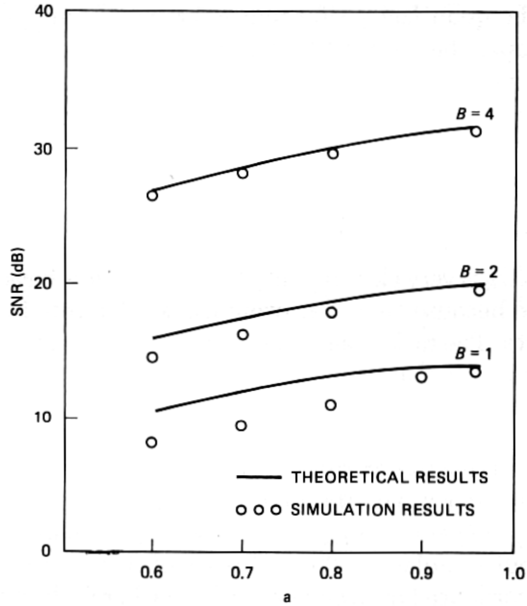

Fig. 2—DPCM performance: SNR vs. value a of the predictor coefficient. Source: Gaussian first-order Markov source with correlation $\rho = 0.9625$. Quantizer: B-bit quantizer optimized for Gaussian signals. Predictor: Previous sample prediction with coefficient a.

The value a_{opt} is not critical, however. We use $a_{opt} \approx \rho$ to determine the total MMSE:

$$\min\{\sigma_t^2\} = \epsilon_q^2 \frac{1 - \rho^2}{1 - \epsilon_q^2 \rho^2} \cdot \sigma_x^2. \tag{18}$$

Due to feedback there is an increase in variance by a factor $(1 - \epsilon_q^2 \rho^2)^{-1}$. Figure 2 shows the dependence of the signal-to-noise ratio (SNR) on the value of the predictor coefficient for various B-bit quantizers and compares theoretical results obtained from eq. (18) with simulation results. It is seen that these are useful approximations in the vicinity of the optimum setting of the predictor coefficient (this is the region where the difference signal is almost white noise, and the simulations reveal that the assumption of a vanishing correlation between quantization error and input signal holds in this case).

In an "ideal" Nth order DPCM system the prediction of an input sample x_n is based on previous input samples $x_{n-j}; j = 1,2 \ldots, N$. Hence the prediction scheme is essentially that of D*PCM (see Fig. 1c), but it is important to realize that these schemes are optimal for different frequency responses $P(\omega)$ of the predictor as will be seen later in this paper. Equation (12) shows that the optimum predictor in DPCM is the MMSE predictor. The prediction error variance is given as

$$\sigma_d^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega)|1 - P(\omega)|^2 d\omega \tag{19}$$

and its minimum is reached if the random sequence $\{d_n\}$ is a white noise sequence of variance η_x^2 (*full-whitening*). Thus we have

$$\min\{\sigma_d^2\} = \eta_x^2, \tag{20}$$

where η_x^2 is the minimum prediction error variance to be obtained from the random sequence $\{x_n\}$ by passing its samples through a prediction error filter with frequency response $1 - P_{opt}(\omega)$ such that

$$S_x(\omega)|1 - P_{opt}(\omega)|^2 = \eta_x^2. \tag{21}$$

We note that this minimum can be obtained for any stationary random sequence if the predictor impulse response is of semi-infinite length, and an Nth order predictor can be employed if the random sequence is Markovian of order N. For a given pds $S_x(\omega)$ Kolmogoroff[20] has given the minimum of the prediction error variance as

$$\eta_x^2 = \min\{\sigma_d^2\} = \exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e S_x(\omega) d\omega\right]. \tag{22}$$

This variance η_x^2 is a positive quantity if the process is undetermined, i.e., if $S_x(\omega)$ is zero at most at a countable set of frequencies. Otherwise the signal is perfectly predictable and the prediction error variance is zero then. The normalized prediction error variance

$$\gamma_x^2 = \frac{\eta_x^2}{\sigma_x^2} = \frac{\exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e S_x(\omega) d\omega\right]}{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) d\omega} \tag{23}$$

is called spectral flatness measure[21] and its inverse is the optimally obtainable prediction gain. The spectral flatness measure can be interpreted as the ratio of the geometric mean of the pds $S_x(\omega)$ to its arithmetic mean. It is easy to show[21] that

$$0 \leqslant \gamma_x^2 \leqslant 1. \tag{24}$$

From eq. (12) we finally find that

$$\min_{\text{DPCM}} \{\sigma_t^2\} = \epsilon_q^2 \cdot \eta_x^2, \tag{25}$$

2.3 D*PCM

The quantization error with pds $S_q(\omega) = \sigma_q^2$ is filtered by the linear decoder network with frequency response $H(\omega) = G^{-1}(\omega)$ (see Fig. 1c).

Thus the total error variance is

$$\sigma_t^2 = \sigma_q^2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega)|^2 d\omega, \tag{26}$$

i.e., the quantization error variance is increased by the power transfer factor

$$\alpha = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega)|^2 d\omega. \tag{27}$$

The variance of the quantization error depends on the variance of the difference signal d^*_n:

$$\sigma_q^2 = \epsilon_q^2 \cdot \sigma_{d^*}^2 = \epsilon_q^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) |G(\omega)|^2 d\omega. \tag{28}$$

Thus the total error variance is

$$\sigma_t^2 = \epsilon_q^2 \cdot \alpha \cdot \sigma_{d^*}^2, \tag{29}$$

i.e., the total error variance of D*PCM is α times that of an equivalent DPCM coding scheme (with $\sigma_d^2 = \sigma_{d^*}^2$ in the case of fine quantizing and identical prediction filters). The impulse response of the postfilter is a sequence $\{1, h_1, h_2, \cdots\}$; thus we have

$$\alpha = 1 + \sum_{k=1}^{\infty} h_k^2 \geqslant 1 \tag{30}$$

and we find that DPCM outperforms D*PCM for *any* choice of the predictor network.[22] This result also implies that the optimum performance of DPCM is better than the optimum performance of D*PCM. This fact will be shown very shortly in a slightly broader context.

Example 2: We calculate now the total MSE of a simple D*PCM scheme which employs a first-order predictor of value a. The prediction error variance has been given in eq. (14). The power transfer factor of the decoder network is

$$\alpha = 1 + a^2 + a^4 + \cdots = (1 - a^2)^{-1}. \tag{31}$$

Thus the total MSE is

$$\sigma_t^2 = \epsilon_q^2 \cdot \alpha \cdot \sigma_{d^*}^2 = \epsilon_q^2 \frac{1 - 2a\rho + a^2}{1 - a^2} \cdot \sigma_x^2. \tag{32}$$

Note that MMSE prediction ($a = \rho$) results in the same MSE as that of standard PCM,[3] and that $\sigma_t^2 \to \infty$ for any ρ if $a \to 1$. The minimum total MSE is obtained for

$$a_{opt} = \frac{1}{\rho}(1 - \sqrt{1 - \rho^2}), \tag{33}$$

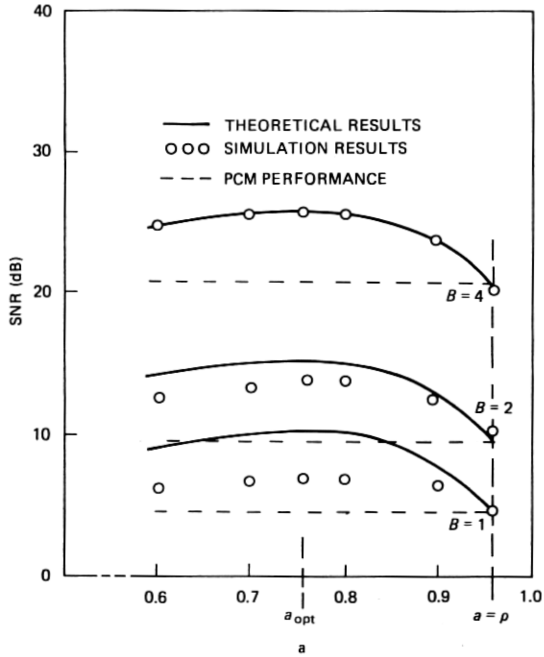

Fig. 3—D*PCM performance: SNR vs. value a of the predictor coefficient. Source and coder same as in Fig. 2.

and its value is

$$\min\{\sigma_t^2\} = \epsilon_q^2 \sqrt{1-\rho^2} \cdot \sigma_x^2. \tag{34}$$

Thus it is possible to reduce the total noise variance by a factor $\sqrt{1-\rho^2}$ instead of $(1-\rho^2)$ as in ideal DPCM (with negligible feedback). It is interesting to note that this is the same reduction that we can obtain with a block quantization scheme of blocklength 2 (this is a scheme where the sum and the difference of adjacent samples are quantized independently).[23] Figure 3 demonstrates this fact that the SNR is lower than that obtainable with DPCM (Fig. 2). In the case of quantizing with one and two bits/sample there is, as expected, a significant difference between theory and measurements especially for a-values which are not close to the correlation coefficient ρ. This difference is now explained by the fact that, in the case of coarse quantization, the white noise assumption [eq. (8)] does not hold for correlated quantizer input samples. Simulations have revealed that higher signal-to-noise ratios can be obtained by choosing a decoder coefficient that is higher than that of the coder. Note also, from Fig. 3, that no gain over PCM is obtainable for the specific case $a = \rho$.

We come back now to the total MSE of D*PCM. Equation (29) is given explicitly as

$$\sigma_t^2 = \epsilon_q^2 \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) |1 - P(\omega)|^2 d\omega \right] \times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} |1 - P(\omega)|^{-2} d\omega \right] \quad (35)$$

where we have used our assumptions $G(\omega) = 1 - P(\omega)$ and $H(\omega) = G^{-1}(\omega)$ (both assumptions will be reviewed in the next section). Note that the term in the first brackets reduces to η_x^2 if $P(\omega)$ is the MMSE predictor [full-whitening, see eq. (21)]. The difference signal d^*_n is white noise then, and, since the last term in eq. (35) is the power transfer factor for white noise inputs, it equals $\sigma_x^2/\sigma_{d^*}^2 = \sigma_x^2/\eta_x^2$. Therefore the total MSE is $\sigma_t^2 = \epsilon_q^2 \cdot \sigma_x^2$, and thus equals that of PCM.[3] Note that this statement is based on assumptions given in the remark in Section 2.2. We relate now the minimum total MSE obtainable with D*PCM to that of DPCM. By applying Schwarz's inequality† to eq. (35) we obtain

$$\min_{\text{D*PCM}} \{\sigma_t^2\} \geqslant \epsilon_q^2 \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{S_x(\omega)}\, d\omega \right]^2 = \epsilon_q^2 \sigma_{\sqrt{x}}^4 \quad (36)$$

where $\sigma_{\sqrt{x}}^2$ defines the variance of a process with pds $\sqrt{S_x(\omega)}$.

Equality is obtained if

$$|1 - P(\omega)|^2 \sqrt{S_x(\omega)} = \text{const.} \quad (37)$$

We conclude that the squared-magnitude frequency response of the D*PCM encoder has to be inversely proportional to the square root of the pds of the input (*half-whitening*). Therefore the optimum frequency response $1 - P(\omega)$ is a MMSE prediction error filter for a pds $\sqrt{S_x(\omega)}$. The corresponding minimum $\eta_{\sqrt{x}}^2$ of the prediction error variance would be just the square root of that to be obtained from an optimum prediction of a process with pds $S_x(\omega)$ [this can be verified from eq. (22)]:

$$|1 - P(\omega)|^2 \sqrt{S_x(\omega)} = \eta_{\sqrt{x}}^2 = \eta_x \quad (38)$$

However, it is important to realize at this point that it is not certain that the equality in eq. (37) can be obtained at all in practice. Indeed, equality can only be expected if $\sqrt{S_x(\omega)}$ happens to be a rational spectrum, since such a spectrum can always be modeled by white noise passed through a purely recursive linear filter with poles inside the unit circle. It is clear that such a process can be whitened by a one-step ahead prediction error filter $1 - P(\omega)$. In general, however, the pds $\sqrt{S_x(\omega)}$ is not rational, and the minimum total error variance to be obtained by D*PCM coding is

† $\int_I f_1^2(x)dx \cdot \int_I f_2^2(x)dx \geqslant [\int_I f_1(x) f_2(x) dx]^2$ with equality if $f_1^2(x)/f_2^2(x) = \text{const.}$ for any square-integrable functions $f_1(\cdot)$ and $f_2(\cdot)$.

greater than the right-hand term of eq. (36). Example 2 has shown that an improvement over PCM can be obtained with D*PCM. By applying Schwarz's inequality to the right-hand term of eq. (36) we have

$$\epsilon_q^2 \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{S_x(\omega)}\, d\omega \right]^2 \leqslant \epsilon_q^2 \sigma_x^2 = \min_{\text{PCM}} \{\sigma_t^2\} \tag{39}$$

with equality only if $\{x_n\}$ is a white noise sequence, i.e., an improvement over PCM can be obtained for all nonwhite processes. Equation (30) has already indicated that DPCM outperforms D*PCM for any nonwhite pds $S_x(\omega)$. We can make this statement more quantitative by comparing the total error variances of D*PCM (in the most favorable case, given if the pds $\sqrt{S_x(\omega)}$ is rational) and that of DPCM. To be more specific, we use eqs. (22)–(25) and (36), and find

$$\frac{\min\limits_{\text{DPCM}} \{\sigma_t^2\}}{\min\limits_{\text{D*PCM}} \{\sigma_t^2\}} \leq \frac{\exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e S_x(\omega) d\omega \right]}{\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{S_x(\omega)}\, d\omega \right]^2}$$

$$= \frac{\eta_x^2}{\sigma_{\sqrt{x}}^4} = \left[\frac{\eta_{\sqrt{x}}^2}{\sigma_{\sqrt{x}}^2} \right]^2 = \gamma_{\sqrt{x}}^4 \leqslant 1. \tag{40}$$

We conclude that the ratio of the total error variances of DPCM and D*PCM is upperbounded by the squared spectral flatness measure $\gamma_{\sqrt{x}}^2$ of the pds $\sqrt{S_x(\omega)}$. The validity of the right-hand inequality in eq. (40) has already been stated in eq. (24). Equality is obtained if and only if $\{x_n\}$ is a white noise sequence. In other words, DPCM outperforms D*PCM for all nonwhite spectra.

Example 3: Example 1 has shown that the total MSE of a DPCM scheme with a first-order predictor is smaller by a factor $\beta^2 = 1 - \rho^2$ than that of PCM (provided that the quantizer performance factors are identical in both cases). In D*PCM, the corresponding factor is $\beta^2 = \sqrt{1 - \rho^2}$ (see example 2). The value ρ is in both examples the one-lag normalized autocorrelation coefficient of an otherwise arbitrary (though stationary) random input sequence. Let us assume now that this sequence is a first-order Markovian sequence with autocorrelation

$$R_x(k) = \rho^{|k|} \cdot \sigma_x^2 \tag{41}$$

and pds

$$S_x(\omega) = \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos \omega} \sigma_x^2. \tag{42}$$

The best DPCM performance is still obtained with a first-order predictor since $\{x_n\}$ is Markovian. The D*PCM reduction factor $\beta^2 = \sqrt{1-\rho^2}$ is not optimal, however. It is interesting to compare it with the upper bound to be obtained without the constraint of the realizability of the prefilter.

Using eqs. (36) and (42) we find

$$\min_{\text{D*PCM}} \{\sigma_t^2\} \geqslant \epsilon_q^2 \cdot \beta^2 \cdot \sigma_x^2 \tag{43}$$

where

$$\beta^2 = (1-\rho^2)\left(\frac{2}{\pi}\right)^2 F^2\left(\frac{\pi}{2}, \rho\right). \tag{44}$$

$F(\pi/2, \rho)$ is the complete elliptical integral of the first kind:

$$F\left(\frac{\pi}{2}, \rho\right) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1-\rho^2 \sin^2\phi}}$$
$$= \frac{\pi}{2}\left[1 + \left(\frac{1}{2}\right)^2 \rho^2 + \left(\frac{1\cdot 3}{2\cdot 4}\right)^2 \rho^4 + \cdots\right]. \tag{45}$$

We conclude that a D*PCM coding of a first-order Markov source is upperbounded by a reduction factor

$$\beta^2 = (1-\rho^2)\left[1 + \frac{1}{4}\rho^2 + \frac{9}{64}\rho^4 + \cdots\right]^2 \tag{46}$$

instead of $(1-\rho^2)$ for DPCM and $\sqrt{1-\rho^2}$ for the one-tap D*PCM. For $\rho = 0.9625$ the one-tap D*PCM coding results in a total MSE that is by a factor 3.7 greater than that of (ideal) DPCM, and it is lower-bounded by a factor 3.0 (for $\rho = 0.85$ the corresponding values are 1.9 and 1.8, respectively).

The foregoing discussion has shown that D*PCM is a suboptimal coding scheme if the performance criterion is the unweighted total error variance. The next section will demonstrate that D*PCM is an optimum coding scheme for a specific frequency-weighted error criterion, and Section IV will show that its performance bounds the overall performance of DPCM coders in the presence of channel errors. As a final remark we mention that D*PCM based on adaptive prediction with a separate transmission of the predictor coefficients has been used recently for speech coding purposes[24] to improve the subjective performance.

III. NOISE-FEEDBACK CODING

In the preceding chapter we have optimized pre- and postfilters for a quantizing scheme (D*PCM) and have shown that DPCM has a superior performance for all nonwhite processes. At a first glance this seems to

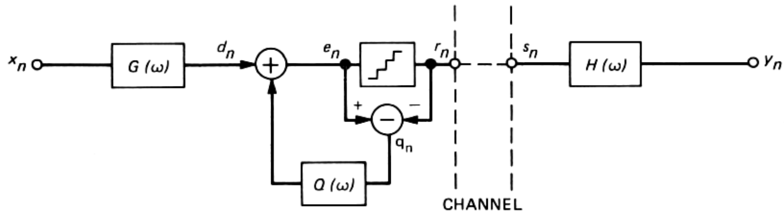

Fig. 4—Noise-feedback coding scheme.

be surprising since D*PCM coding is also based on linear filtering. In the D*PCM analysis, however, we did not take into account that not only the pre- and postfilters are under our control, but also the quantizer. To be more specific, we may additionally filter the quantization noise whereas in typical pre- and postfiltering applications it is the channel that is perturbed by noise; that noise is not under control of the designer. Figure 4 demonstrates how the filtering of the quantization noise is obtained. The quantization noise, i.e., the difference between input and output of the quantizer, is fed back through a linear filter with frequency response $Q(\omega)$ and is added to the input. $Q(\omega)$ is required to have a minimum delay of one sampling time for stability reasons. The purpose of the feedback scheme is a reshaping of the spectrum of the quantization noise such that the total error variance is minimum. It should also be mentioned at this point that there exists also another linearly equivalent scheme, the direct-feedback coder, which has been studied by Brainard and Candy.[15] It uses a prefilter of frequency response $A(\omega)$, and the output of the quantizer (not the quantization noise!) is first passed through a feedback filter of frequency response $B(\omega)$ and then added to the prefiltered signal to form the quantizer input. The equivalence to the noise-feedback coder is given by $G(\omega) = A(\omega)/(1 - B(\omega))$ and $1 - Q(\omega) = 1/(1 - B(\omega))$.

3.1 Derivation of the basic formula

Let us, as before, represent the quantizer as a device that adds signal-independent white noise of pds $S_q(\omega) = \sigma_q^2$ to the signal. Due to this assumption we can also replace the feedback-quantizer with a nonwhite noise source of pds

$$\begin{aligned} S_n(\omega) &= S_q(\omega)|1 - Q(\omega)|^2 \\ &= \sigma_q^2|1 - Q(\omega)|^2. \end{aligned} \tag{47}$$

The feedback acts as linear filter on the open-loop quantizing noise, and the effective quantization noise is colored noise then. Additionally we may introduce a subjective noise-weighting function $S_w(\omega)$ whose inverse describes the sensitivity of the sink to uncorrelated noise. A small value of $S_w(\omega)$ indicates that a high error variance is acceptable in that specific frequency region and vice versa. It is well known that such a frequency-weighting can only serve as a first approximation of noise sensitivity since it does not take into account nonlinear effects and dependencies on local properties of the signals. Nevertheless, the frequency-weighted noise power is a quite popular design criterion partly due to the lack of better distortion measures. What is worth emphasizing is that $S_w(\omega)$ can be interpreted as the squared-magnitude frequency response of a noise-weighting filter. Therefore it is reasonable to assume that $S_w(\omega)$ is a rational function of ω. $S_w(\omega)$ can also be explained as the output pds of such a weighting filter whose input is a white noise process. This interpretation suggests to define a variance

$$\sigma_w^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_w(\omega) d\omega, \tag{48}$$

a minimum prediction error variance

$$\eta_w^2 = \exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e S_w(\omega) d\omega \right], \tag{49}$$

and a spectral flatness measure

$$\gamma_w^2 = \eta_w^2 / \sigma_w^2 \tag{50}$$

in accordance with eqs. (7), (22), and (23).

The MSE of a NFC coder is given by

$$\sigma_t^2 = \epsilon_q^2 \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) |G(\omega)|^2 d\omega \right] \cdot \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_w(\omega) |H(\omega)|^2 |1 - Q(\omega)|^2 d\omega \right] \tag{51}$$

where we have used eqs. (26), (28), and (47). Note that the MSE does not include a linear distortion term resulting from a possible mismatch between prefilter and postfilter. We are now free to choose linear filters $G(\omega)$, $H(\omega)$, and $Q(\omega)$ such that the frequency-weighted error variance is minimized. This is obtained by a proper preshaping of the signal spectrum *and* the quantization noise prior to transmission. The general design of this NFC scheme has been studied by Kimme and Kuo in the context of picture coding.[14] Our objective is to show the connection of this noise-feedback scheme with those discussed so far and to show how subjective noise-weighting functions influence the design. Our approach will also be different from that in Ref. 14 since a constraint is not needed in the optimization procedure. The MMSE design problem is to find the optimum combination of prefilter $G(\omega)$, postfilter $H(\omega)$, and feedback filter $Q(\omega)$ for a given pds $S_x(\omega)$ of the input signal, and a weighting function $S_w(\omega)$. Let us first assume that $G(\omega)$ and $Q(\omega)$ are given and let us determine $H(\omega)$. In an optimized system the total error $x_n - y_n$ must be orthogonal to the output of the postfilter and thus orthogonal to the data r_n used in that filter (the sequence $\{r_n\}$ is the decoder input sequence). It can be shown that this condition also holds in the case of weighted total errors. If the postfilter $H(\omega)$ is not constrained to be physically realizable (its characteristics can always be approximated arbitrarily closely by allowing for a sufficient time delay), the optimum filter is given by the Wiener-Hopf condition

$$H_{opt}(\omega) = \frac{S_{rx}(\omega)}{S_r(\omega)} \tag{52}$$

where $S_{rx}(\omega) = G^*(\omega) \cdot S_x(\omega)$ is the cross-spectrum between the sequences $\{r_n\}$ and $\{x_n\}$. Thus we see that

$$H_{opt}(\omega) = \frac{G^*(\omega) \cdot S_x(\omega)}{|G(\omega)|^2 S_x(\omega) + S_q(\omega)|1 - Q(\omega)|^2}. \tag{53}$$

We shall restrict our attention to the case of a small quantization noise variance:

$$S_q(\omega) \ll \frac{|G(\omega)|^2 S_x(\omega)}{|1 - Q(\omega)|^2} \quad \text{for } all\ \omega. \tag{54}$$

Equation (53) becomes

$$H_{opt}(\omega) = \frac{1}{G(\omega)}, \tag{55}$$

so that $G(\omega)$ and $H(\omega)$ are reciprocal filters for *any* given $G(\omega)$ which does not violate the assumption of eq. (54). It is seen that choosing reciprocal coding and decoding filters is not just a convenience but a requirement by the MSE criterion. By applying Schwarz's inequality to eq. (51), we find

$$\min\{\sigma_t^2\} = \epsilon_q^2 \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{S_x(\omega) S_w(\omega)} |1 - Q(\omega)|^2 \, d\omega \right]^2 \tag{56}$$

for any given $S_w(\omega)$ and $Q(\omega)$. This minimum is reached if

$$|G_{opt}(\omega)|^2 = C^2 \sqrt{\frac{S_w(\omega)|1 - Q(\omega)|^2}{S_x(\omega)}} \tag{57}$$

where C is a constant. Equations (51) and (56) can be used to calculate the error variances of various coding schemes.

3.2 Optimization of the noise-feedback coder

We shall now derive two conditions which have to be met by $Q(\omega)$ and $G_{opt}(\omega)$, respectively. We apply again Schwarz's inequality to eq. (56)

and thus have

$$\min\{\sigma_t^2\} \leqslant \epsilon_q^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) S_w(\omega) |1 - Q(\omega)|^2 d\omega \tag{58}$$

with equality if the integrand is a constant. Note that the term $1 - Q(\omega)$ describes a prediction error structure. Therefore equality is obtained by choosing $Q(\omega)$ to be the MMSE predictor of a random sequence of pds $S_x(\omega) \cdot S_w(\omega)$ and we have [in the notation of eq. (21)]:

$$S_x(\omega) S_w(\omega) |1 - Q_{opt}(\omega)|^2 = \eta_{xw}^2 = \eta_x^2 \cdot \eta_w^2 \tag{59}$$

where $Q_{opt}(\omega)$ is the MMSE predictor that whitens a random sequence of pds $S_x(\omega) \cdot S_w(\omega)$, and where $\eta_x^2 \eta_w^2$ is its MMSE. The right-hand equality in eq. (59) can be obtained from Kolmogoroff's result [eq. (22)] by substituting $S_x(\omega)$ with its frequency-weighted version $S_x(\omega) S_w(\omega)$. When comparing eqs. (56) and (59) we find

$$\min_{\text{NFC}} \{\sigma_t^2\} = \epsilon_q^2 \cdot \eta_x^2 \cdot \eta_w^2. \tag{60}$$

We conclude that the frequency-weighted total error variance of an NFC scheme with given quantizer is determined by the product of the prediction error variances of the spectra $S_x(\omega)$ and $S_w(\omega)$, and that $Q_{opt}(\omega)$ is the optimal predictor of a pds $S_x(\omega) \cdot S_w(\omega)$. We also have, from eqs. (57) and (59), that $S_x(\omega) |G_{opt}(\omega)|^2 = \text{const.}$ This implies, however, that $|G_{opt}(\omega)|^2$ is a filter which whitens $S_x(\omega)$ which has been assumed to be rational. Thus a sufficient condition for an optimum NFC scheme is to choose as a prefilter an MMSE optimized prediction error filter:

$$G_{opt}(\omega) = 1 - P_{opt}(\omega) \tag{61}$$

where $P_{opt}(\omega)$ is defined by

$$S_x(\omega) |1 - P_{opt}(\omega)|^2 = \eta_x^2. \tag{62}$$

Note the important fact that the overall performance is optimized if the prefilter is an MMSE prediction-error filter that whitens the input process of pds $S_x(\omega)$ and that this result holds for any choice of the weighting function. It turns out that we have to modify the quantization noise feedback loop but not the prefilter if a weighting of the noise has to be taken into account. We finally find from eqs. (20), (47), (59), and (62) that the weighted error spectrum $S_t(\omega) = \epsilon_q^2 \sigma_d^2 S_w(\omega) |1 - Q_{opt}(\omega)|^2 \cdot |1 - P_{opt}(\omega)|^{-2}$ is constant and thus equals the total error variance [see eq. (60)]. We finally note that these optimization results can also be extracted from the Kimme and Kuo paper.[14] Musmann has recently derived equivalent results based on a more information-theoretical analysis.[25]

We shall now briefly discuss two special cases.

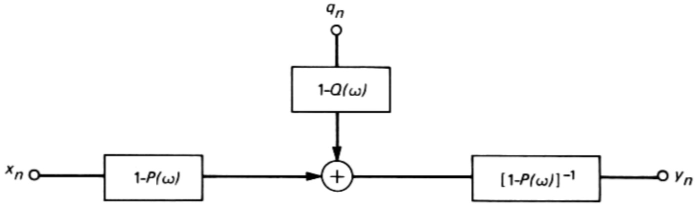

Fig. 5—Model of NFC coder with prediction error filter as input.

Special case: $S_w(\omega) = 1$ for all ω (DPCM). A comparison of eqs. (59) and (62) reveals that $Q_{opt}(\omega) = P_{opt}(\omega)$, i.e., the predictor prefilter and the feedback filter have to have identical frequency responses. It is easy to see[26] that these filters can then be combined to yield the DPCM structure of Fig. 1b. A DPCM structure is indeed defined by having $Q(\omega) = P(\omega)$ and Fig. 5, which is a model of a noise-feedback coder, reveals that no frequency-weighting of the quantization noise is possible in this case for *any* choice of the predictor (provided that pre- and postfilters are reciprocals), since the quantization noise passes both the feedback filter with frequency response $1 - P(\omega)$ and its inverse, the postfilter.

Special case: $S_w(\omega) \propto S_x^{-1}(\omega)$ for all ω (D*PCM). A weighting function that is in some sense inverse to the pds of the signal is of importance for quantizing acoustic signals since it may avoid a masking of weak signal energies in specific frequency ranges by the quantization noise.

For $S_w(\omega) \propto S_x^{-1}(\omega)$ we find from eq. (59) that $|1 - Q_{opt}(\omega)|^2$ has to be a constant. This clearly means that $Q_{opt}(\omega) = 0$ for all ω; i.e., the best coding scheme is now D*PCM. The prefilter is, of course, as for all optimal NFC coders, a whitening filter. The D*PCM scheme is suboptimum if it is used in connection with other weighting functions. The special case of $S_w(\omega) = 1$ has been discussed in detail in Section II; we have seen there that for $S_w(\omega) = 1$ and a MMSE predictor $P(\omega)$ no reduction in total error variance over PCM is obtainable. The above discussion reveals that the same coder is, however, optimal for the specific noise-weighting function $S_w(\omega) \propto S_x^{-1}(\omega)$. Indeed, subjective gains of about 6–10 dB have been reported for speech signals, for this choice of the prefilter.[26,27]

Table I lists various NFC configurations. The performance of some suboptimal coders will be compared with the optimal NFC scheme in the next part of this section. An elementary example will suggest the manner in which the NFC design influences the total error variance.

Example 4: Assume a noise-feedback coder with just one tap, i.e., $Q(\omega) = q{\cdot}\exp(-j\omega)$, and a prefilter with the structure of a one-tap prediction-error filter: $G(\omega) = 1 - a{\cdot}\exp(-j\omega)$ (see Fig. 6). Further assume a sequence with an adjacent-sample correlation ρ. We have $H(\omega) = G^{-1}(\omega)$

Table I — Performance comparison of NFC coder configurations

Noise weighting	NFC	DPCM $Q(\omega) = P(\omega)$	D*PCM $Q(\omega) = 0$
$S_w(\omega) = 1$ (no weighting)	$\hat{=}$ DPCM	Optimal coder: $\lvert 1 - P \rvert^2 \propto S_x^{-1}$	Suboptimal coder (upper bound half-whitening): $\lvert 1 - P \rvert^2 \propto S_x^{-1/2}$
$S_w(\omega) \propto S_x^{-1}(\omega)$	$\hat{=}$ D*PCM	Suboptimal coder	Optimal coder (full-whitening): $\lvert G \rvert^2 = \lvert 1 - P \rvert^2 \propto S_x^{-1}$
All other cases:	Optimal coder	Suboptimal coder	Suboptimal coder

and the unweighted total MSE $[S_w(\omega) = 1]$ can be derived from eq. (51); we shall omit the intermediate steps. The final result is

$$\sigma_t^2 = \epsilon_q^2 \frac{1 + a^2 - 2a\rho}{1 - a^2} (1 + q^2 - 2aq) \cdot \sigma_x^2$$

$$a < 1;\ q \text{ arbitrary.} \tag{63}$$

For $a = 0$ and $q = 0$ we have the PCM result of eq. (10). For $a = 0$ and finite q we obtain $\sigma_t^2 = \epsilon_q^2(1 + q^2) \cdot \sigma_x^2$, i.e., noise feedback without prefiltering increases the MSE by a factor $1 + q^2$ over that of PCM. For $q = 0$ and finite $a < 1$ we have the case of D*PCM, i.e., prefiltering followed by quantization [eq. (32)]. The best choice of q is $q = a$ if a is given; the scheme reduces then to DPCM [eq. (1)] and reaches its MMSE for $a = q = \rho$. The different cases have been tabulated in Table II.

3.3 Suboptimal coding schemes

The last section has shown that noise-feedback coding is a scheme that allows for the optimal shaping of the spectra of the input signal *and* the quantization noise such that the noise-weighted overall error variance is minimized. The prefilter is in all cases a MMSE prediction error filter $1 - P_{opt}(\omega)$ which performs a decorrelation of the input signal. The optimal scheme reduces to DPCM ($Q_{opt}(\omega) = P_{opt}(\omega)$) in the case of unweighted noise ($S_w(\omega) = 1$), and to D*PCM ($Q_{opt}(\omega) = 0$) in the case of $S_w(\omega) \propto S_x^{-1}(\omega)$. In all other cases, only the general NFC scheme (with $Q(\omega) \neq P(\omega)$) is optimal.

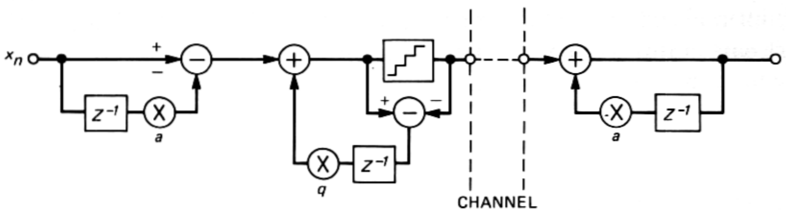

Fig. 6—One-tap NFC coder of Example 4.

Table II—Error variances of coding schemes with one predictor coefficient a_{opt} and one feedback coefficient q_{opt}

	$\min\{\sigma_t^2\}$	q_{opt}	a_{opt}
PCM	$\epsilon_q^2 \cdot \sigma_x^2$	0	0
Noise-feedback PCM	$\epsilon_q^2(1 + q^2)\sigma_x^2$	$\neq 0$	0
D*PCM	$\epsilon_q^2 \sqrt{1-\rho^2}\, \sigma_x^2$	0	$\frac{1}{\rho}(1 - \sqrt{1-\rho^2})$
DPCM (ideal)	$\epsilon_q^2 (1 - \rho^2)\sigma_x^2$	ρ	ρ
DPCM (real, i.e., with coarse quantization)	$\epsilon_q^2 \frac{1-\rho^2}{1-\epsilon_q^2\rho^2} \sigma_x^2$	a_{opt}	$<\rho$ [see eq. (17)]

This section compares the performances of various suboptimal coding schemes which belong to the class of NFC schemes and which are derived by choosing suboptimum prefilters and feedback filters. The corresponding error variances can be derived directly from eq. (51). We shall omit the intermediate steps in the calculations and optimizations of σ_t^2 and shall present the final results only.

PCM. PCM results if $G(\omega) = H(\omega) = 1$ and $Q(\omega) = 0$. We then have

$$\min_{\text{PCM}} \{\sigma_t^2\} = \epsilon_q^2 \sigma_x^2 \sigma_w^2. \tag{64}$$

The quantity σ_w^2 as defined in eq. (48) represents the subjective gain if white quantization noise is weighted.

Noise-Feedback PCM. We have $G(\omega) = H(\omega) = 1$, and $Q_{opt}(\omega)$ is the MMSE predictor of the weighting function $S_w(\omega)$. Hence we have

$$\min_{\substack{\text{noise-feedback}\\ \text{PCM}}} \{\sigma_t^2\} = \epsilon_q^2 \sigma_x^2 \eta_w^2. \tag{65}$$

This result shows that η_w^2 as defined in eq. (49) represents the error reduction obtainable if the quantization noise is optimally shaped. Note that optimal noise-shaping reduces the weighted error variance in relation to PCM by a factor $\gamma_w^2 = \eta_w^2/\sigma_w^2$ which is the spectral flatness measure of the weighting function $S_w(\omega)$.

Remark: We have to mention at this point that eq. (65) only holds if γ_w^2 is not close to zero. This implies that the predictibility of $S_w(\omega)$ (which is bounded by the dynamic range of the spectrum[28]) should not be too high. Otherwise one could obviously achieve large reductions in error (including unbounded ones for weighting functions which are zero over a finite segment of the frequency axis; see eq. (22) and the following discussion thereof). The noise reductions are obtained by shifting the noise in frequency to a range where it is less heavily weighted by $S_w(\omega)$. However, the noise-shaping increases the total variance of the quantizer input signal since the correlated quantization noise is added to the input signal. Therefore the quantizer step sizes have to be readjusted accordingly. As a consequence we obtain an increase in the quantizer performance factor ϵ_q^2 and subsequently an increase in total error variance. Spang and Schultheiss[29] have discussed in detail the stability problems involved. They have been interested in noise reductions in oversampled systems where the noise can be shifted into the high-frequency range and subsequently eliminated by lowpass filtering. In that application we have a weighting function that is zero over a finite segment of the frequency axis. As a means for reducing the stability problems Spang and Schultheiss propose to keep the number of feedback elements R finite. In this case the total error variance is given as

$$\sigma_t^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_w(\omega) \left| 1 - \sum_{k=1}^{R} q_k e^{-jk\omega} \right|^2 d\omega. \tag{66}$$

The feedback network can be viewed as an Rth order predictor. Its optimal coefficients q_k; $k = 1,2,\ldots,R$ can be derived from the set of R normal equations whose coefficients are just the Fourier coefficients of $S_w(\omega)$.[30]

D*PCM. We have $G(\omega) = 1 - P(\omega)$ and $Q(\omega) = 0$.

(*i*) *Lower bound* (*half-whitening*). The lower bound is reached if $|1 - P_{opt}(\omega)|^2 = \sqrt{S_w(\omega)/S_x(\omega)}$. We obtain

$$\min_{\text{D*PCM}} \{\sigma_t^2\} = \epsilon_q^2 \sigma_{\sqrt{xw}}^4 \tag{67}$$

where

$$\sigma_{\sqrt{xw}}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{S_x(\omega) \cdot S_w(\omega)}\, d\omega. \tag{68}$$

(*ii*) *Full-whitening*. Let $P(\omega)$ be the MMSE predictor for $S_x(\omega)$. It follows that

$$\sigma_t^2 = \epsilon_q^2 \sigma_{xw}^2 \tag{69}$$

where

$$\sigma_{xw}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) \cdot S_w(\omega) d\omega. \tag{70}$$

DPCM. Let $P(\omega) = Q(\omega)$ be the MMSE predictor for $S_x(\omega)$. It follows that

$$\min_{\text{DPCM}} \{\sigma_t^2\} = \epsilon_q^2 \eta_x^2 \sigma_w^2 \tag{71}$$

Note that optimal prediction has reduced the weighted error variance in relation to PCM by a factor $\gamma_x^2 = \eta_x^2/\sigma_x^2$, i.e., by the spectral flatness measure of $S_x(\omega)$.

Prefiltered DPCM. Let $Q(\omega)$ be the MMSE predictor for $S_x(\omega)$. DPCM results if $P(\omega) = Q(\omega)$. Prefiltered DPCM results if $P(\omega) \neq Q(\omega)$. By varying $P(\omega)$ we obtain

$$\min_{\substack{\text{Prefiltered}\\ \text{DPCM}}} \{\sigma_t^2\} = \epsilon_q^2 \eta_x^2 \sigma_{\sqrt{w}}^4 \tag{72}$$

where

$$\sigma_{\sqrt{w}}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{S_w(\omega)}\, d\omega. \tag{73}$$

From Schwarz's inequality we have $\sigma_{\sqrt{w}}^4 \leqslant \sigma_w^2$. Hence we find the result that the performance of DPCM can be improved by employing an additional prefilter. It is worth emphasizing, however, that we have set $Q(\omega)$ to be the MMSE predictor for $S_x(\omega)$. The optimal NFC scheme results if we are also free to optimize this feedback filter (see below).

NFC. The minimum total error variance of NFC schemes has already been given in Section 3.2 and is repeated here for completeness:

$$\min_{\text{NFC}} \{\sigma_t^2\} = \epsilon_q^2 \eta_x^2 \eta_w^2. \tag{60}$$

Note that optimal prediction *and* optimal noise shaping properties are provided by NFC schemes. Accordingly the total error variance is reduced in relation to PCM by a factor $\gamma_x^2 \cdot \gamma_w^2$ which is the product of the spectral flatness measures of input spectrum and weighting function. We shall see very shortly that this scheme is very close to theoretical bounds in the case of Gaussian input sequences.

Example 5: We shall now evaluate the above derived results for the specific example of a first-order Markov source of variance σ_x^2 and normalized adjacent-sample correlation $\rho \geqslant 0$ whose power density spectrum has already been given in Example 3. Such a source is a useful first approximation for modelling the statistics of speech (with $\rho = 0.85$ in this example) and of television signals (with $\rho = 0.9625$ in this example).

Speech signals. We assume a weighting function which is inversely proportional to the pds of the signal. We have already mentioned in Section 3.2 that subjective gains of about 6–10 dB have been reported for this specific weighting. The foregoing analysis has also revealed that the optimal scheme, i.e. NFC is identical to D*PCM. We have

$$S_w(\omega) = c^2 \cdot S_x^{-1}(\omega) \tag{74}$$

where c^2 can easily be determined if $S_w(\omega)$ is normalized such that

$$\max_{\omega} \{S_w(\omega)\} = 1 \tag{75}$$

Table III — Comparison of coder design parameters

	$S_w(\omega) = 1$	$S_w(\omega) \propto S_x^{-1}(\omega)$
η_x^2	$(1-\rho^2)\cdot\sigma_x^2$	$(1-\rho^2)\cdot\sigma_x^2$
η_w^2	1	$\dfrac{1}{(1+\rho)^2}$
σ_w^2	1	$\dfrac{1+\rho^2}{(1+\rho)^2}$
σ_{xw}^2	σ_x^2	$\eta_x^2\cdot\eta_w^2$
$\sigma_{\sqrt{xw}}^2$	$\dfrac{2}{\pi}\eta_x F\left(\dfrac{\pi}{2},\rho\right)$	$\dfrac{\eta_x}{1+\rho} = \sigma_{xw}$
$\sigma_{\sqrt{w}}^2$	1	$\dfrac{2}{\pi}E\left(\dfrac{\pi}{2},k\right)$

and if, in addition, use is made of the following result:

$$\min_{\omega}\{S_x(\omega)\} = \frac{1-\rho}{1+\rho}\,\sigma_x^2. \tag{76}$$

Table III lists the important quantities for $S_w(\omega) = 1$ and $S_w(\omega) \propto S_x^{-1}(\omega)$. $F(\cdot,\cdot)$ is the normal elliptic integral of the first kind [see eq. (45)], and $E(\cdot,\cdot)$ is of the second kind. We have

$$E\left(\frac{\pi}{2},k\right) = \frac{\pi}{2}\left(1 - \frac{1}{4}k^2 - \frac{3}{64}k^4 - \dots\right). \tag{77}$$

The quantity k is given by

$$k = \sqrt{4\rho}/(1+\rho) \tag{78}$$

and $E(\pi/2, k)$ is close to unity for $\rho \geqslant 0.85$.

From Table III we find that the total error variance can be reduced by a factor $\gamma_x^2 = \eta_x^2/\sigma_x^2 = 1-\rho^2$ by means of optimal prediction and additionally by a factor $\gamma_w^2 = \eta_w^2/\sigma_w^2 = (1+\rho)^{-2}$ by means of optimal noise shaping. For $\rho = 0.85$ the obtainable improvements in weighted signal-to-noise ratio are 5.6 dB and 2.4 dB, respectively. It is interesting to note that a noise-shaping improvement of about 1 dB has been calculated in Refs. 25 and 31 on the basis of experimentally determined speech spectra and noise-weighting functions. The total improvement is 8.0 dB for D*PCM or NFC, and it is 6.5 dB for prefiltered DPCM.

The achievable error variances of various coding schemes can be determined by using eqs. (60) and (64)–(73):

PCM: $$\min\{\sigma_t^2\} = \epsilon_q^2\,\frac{1+\rho^2}{(1+\rho)^2}\,\sigma_x^2 \tag{79a}$$

Noise-feedback PCM: $\min\{\sigma_t^2\} = \epsilon_q^2 \frac{1}{(1+\rho)^2} \sigma_x^2$ (79b)

D*PCM and NFC: $\min\{\sigma_t^2\} = \epsilon_q^2 \frac{1-\rho}{1+\rho} \sigma_x^2$ (79c)

DPCM: $\min\{\sigma_t^2\} = \epsilon_q^2 \frac{(1-\rho)(1+\rho^2)}{(1+\rho)} \sigma_x^2$ (79d)

Prefiltered DPCM: $\min\{\sigma_t^2\} \approx \epsilon_q^2 (1-\rho^2) \left(\frac{2}{\pi}\right)^2 \sigma_x^2.$ (79e)

Television signals. An average video spectrum is flat for frequencies below the line rate and falls at about 6 dB per octave through the rest of the band. A weighting function for such a signal has a negative slope of about 3 dB per octave.[7,31,32] We assume a weighting

$$S_w(\omega) = c^2 \sqrt{S_x(\omega)} \tag{80}$$

and it is not difficult to show that the reduction factor for noise shaping is

$$\gamma_w^2 = \frac{\pi}{2F\left(\frac{\pi}{2}, \rho\right)} = \frac{1}{1 + \frac{1}{4}\rho^2 + \frac{9}{64}\rho^4 + \dots} \tag{81}$$

for this choice of the weighting function. Using $\rho = 0.9625$, we find a performance improvement of 11.3 dB obtainable by optimal prediction and a noise-shaping gain of 2.4 dB. This latter figure is in good agreement with the results of calculations based on experimental data.[25,31]

3.4 Absolute performance bounds

In the foregoing sections we have optimized various coding schemes whose structures had been given beforehand. It is useful to compare the performance of this class of encoders-decoders with absolute performance bounds by consulting the distortion-rate function.[12] A detailed discussion of these bounds for speech and television signals has been given by O'Neal.[31] In the following we shall restrict our attention to source coding of stationary ergodic Gaussian processes. For a fixed rate R there exists a minimum possible average distortion D which is a lower bound for any coding scheme. As above we adopt a frequency-weighted mean-squared error as our distortion measure. D and R are then related parametrically as follows:[33,34]

$$D(\varphi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\{\varphi, S_w(\omega) \cdot S_x(\omega)\} d\omega \tag{82}$$

$$R(\varphi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max\left\{0, \log_2 \frac{S_w(\omega) \cdot S_x(\omega)}{\varphi}\right\} d\omega. \tag{83}$$

Next we define the weighted error spectrum by

$$D = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_t(\omega) d\omega \tag{84}$$

and, by comparing it with eq. (82), we find

$$S_t(\omega) = \begin{cases} \varphi & \varphi \leqslant S_w(\omega) \cdot S_x(\omega) \\ S_w(\omega) S_x(\omega) & \text{otherwise.} \end{cases} \tag{85}$$

Over the frequency range where $\varphi > S_w(\omega) \cdot S_x(\omega)$ we have $S_t(\omega) = S_w(\omega) \cdot S_x(\omega)$. This implies that no signal has to be transmitted over this frequency range since such a measure produces just this error spectrum. Over the frequency range where $\varphi \leqslant S_w(\omega) \cdot S_x(\omega)$ we have $S_t(\omega) = \varphi$, i.e., the weighted error spectrum must be constant. Section 3.2 has shown that this requirement is met by the NFC scheme. In this latter case of small distortions we have $D(\varphi) = \varphi = D$ and, by combining eqs. (22), (49), and (83), we obtain

$$D = 2^{-2R} \cdot \eta_x^2 \cdot \eta_w^2. \tag{86}$$

A comparison with eq. (60) indicates that optimal NFC coding is very close to the distortion-rate bound D. The difference is in the first right-hand term of these equations because $\epsilon_q^2 > 2^{-2R}$ for single-letter quantizers (see Table IV). We finally note that the three right-hand terms in eq. (86) correspond to the terms T_B, T_P, and T_S in O'Neal's paper.[31]

IV. TRANSMISSION ERRORS

Noise-feedback coding schemes (including D*PCM and DPCM) are affected differently from PCM systems by bit errors on the communication channel because the decoder loop causes an error propagation while a PCM error does not propagate in time. The objective of this section is to show the effects of transmission errors in predictive coding systems using some of the results of our above analysis. We shall concentrate on two coding schemes, D*PCM and DPCM, and we shall only use the unweighted mean-squared error criterion. Recall that DPCM and NFC are identical in this case.

4.1 PCM

Let us assume that quantizing noise q_n and channel noise c_n can be modelled as additive noise sources. Thus the total error is

$$t_n = q_n + c_n \tag{87}$$

and its variance is

$$\sigma_t^2 = \sigma_q^2 + \sigma_c^2 + 2E[q_n \cdot c_n]. \tag{88}$$

Table IV — Quantization error variances ϵ_q^2 (Max-quantizers) and channel coefficients γ

	ϵ_q^2		γ	
	1 bit	2 bit	1 bit	2 bit
Uniform pdf	0.25	0.0063	3.0	3.75
Gaussian pdf	0.363	0.118	2.55	4.65
Laplacian pdf	0.5	0.176	2.0	5.3
Gamma pdf	0.667	0.232	1.33	6.28

Totty and Clark have shown[35] that channel errors and quantization errors are uncorrelated if the quantizer structure is that of Max.[36] These quantizers minimize the variance of the quantization noise but not necessarily that of the total error. This approach is of interest if a coding scheme has to operate on noisy channels with small bit-error probabilities which additionally are unknown or changing.† It is also justified by our observation that the step-size of quantizers with a low number of levels is not critical. The channel error variance depends on the bit-error probability P, on the density function of the signal being quantized, and on σ_x^2 (because the input variance determines the quantizer step-size scaling). Thus we have

$$\min_{\text{PCM}} \{\sigma_c^2\} = \epsilon_c^2 \cdot \sigma_x^2 \tag{89}$$

and the normalized channel error variance can be written as

$$\epsilon_c^2 = \gamma \cdot P \tag{90}$$

provided that the codewords are only affected by single bit errors. The channel coefficients γ can be derived following an approach in Ref. 39. Table IV lists these values for 1-bit and 2-bit quantizers.[19] In the case of 1 bit, the channel coefficient is simply given as

$$\gamma = 4 \cdot (1 - \epsilon_q^2). \tag{91}$$

In Table IV the γ-values for 2-bit quantizers are given for the folded binary code with the exception of the uniform probability density function whose γ-value is lower in the case of a natural binary code and is given by:[40]

$$\epsilon_c^2 = 4 \cdot P(1 - \epsilon_q^2) \tag{92}$$

where

$$\epsilon_q^2 = 2^{-2B} \tag{93}$$

with B as the number of bits.

† A re-optimization of quantizers for noisy channels has been discussed in Refs. 37 and 38.

The total error variance can thus be calculated as

$$\min_{\text{PCM}} \{\sigma_t^2\} = (\epsilon_q^2 + \epsilon_c^2) \cdot \sigma_x^2 = (\epsilon_q^2 + \gamma \cdot P) \cdot \sigma_x^2 \tag{94}$$

on the assumption of a vanishing correlation between the two errors.

4.2 Noise-feedback coding

The analysis of the predictive quantizing systems in the presence of channel errors is essentially identical to that of the last section if we substitute the nonwhite noise source $S_n(\omega)$ in eq. (47) with

$$\begin{aligned} S_n(\omega) &= \sigma_q^2 |1 - Q(\omega)|^2 + \sigma_c^2 \\ &= [\epsilon_q^2 |1 - Q(\omega)|^2 + \epsilon_c^2] \cdot \sigma_d^2 \end{aligned} \tag{95}$$

and if the assumption of eq. (54) still holds. Thus it is possible to reoptimize the various coders by following the same procedure as in the last section. We shall concentrate in this section on two coding schemes, D*PCM and DPCM. The objective here is to show that the D*PCM performance provides a bound for all predictive quantizing schemes if the channel is noisy.

For both schemes the contribution of the channel errors σ_c^2 on the total error variance can directly be derived from the D*PCM results of Section 2.3 by replacing ϵ_q^2 with ϵ_c^2. We then have, from eq. (29),

$$\begin{aligned} \text{D*PCM: } \sigma_c^2 &= \epsilon_c^2 \cdot \alpha \cdot \sigma_{d*}^2 \\ \text{DPCM: } \sigma_c^2 &= \epsilon_c^2 \cdot \alpha \cdot \sigma_d^2 \end{aligned} \tag{96}$$

where α is the power transfer factor of eq. (27). Therefore the variance of white noise on the channel is increased in the decoder network by a factor $\alpha \geqslant 1$. This error accumulation does not imply that the effect of transmission errors in D*PCM and DPCM is more severe than in PCM, because the *generated* noise variance depends now on the variance of the difference signal and can thus be influenced by the prefilter. Gains over PCM even for noisy channels have indeed been reported recently.[41,42] The discussion of Section 2.3 has already shown that the total MSE can be smaller than that of PCM; it has also shown that coder and decoder should have inverse networks.

4.2.1 D*PCM

Transmission errors contribute to the total error in exactly the same way as quantization noise. The total error

$$t_n = (q_n + c_n) * h_n \tag{97}$$

has a variance

$$\sigma_t^2 = (\epsilon_q^2 + \epsilon_c^2) \cdot \alpha \cdot \sigma_{d*}^2. \tag{98}$$

An optimized D*PCM scheme minimizes at the same time both error contributions. The mean-squared errors are those given in Section 2.3 if ϵ_q^2 is replaced with $\epsilon_q^2 + \epsilon_c^2$. Note that the optimal filters in D*PCM schemes do *not* depend on the bit-error rate. We have seen that half-whitening of the input spectrum provides a D*PCM performance bound. From eq. (36) we conclude that the minimum channel error variance is given as

$$\min_{\text{D*PCM}} \{\sigma_c^2\} = \epsilon_c^2 \sigma_{\sqrt{x}}^4. \tag{99}$$

4.2.2 DPCM

Section 3.2 has revealed that D*PCM has the same total quantization error variance as PCM if the input is full-whitened (MMSE prediction). The D*PCM postfilter is then identical with that of DPCM if this latter scheme has been optimized for the noiseless channel. These observations imply that transmission errors cause the same channel error variances in DPCM as in PCM if an MMSE predictor is employed. Smaller MSE values, i.e., improvements over PCM, can be gained by *reducing* the whitening effect. The quantization noise MSE, however, increases then. The total error

$$t_n = q_n + c_n * h_n \tag{100}$$

has a variance

$$\sigma_t^2 = (\epsilon_q^2 + \alpha \epsilon_c^2)\sigma_d^2. \tag{101}$$

In the case of high bit-error probabilities ($\alpha\epsilon_c^2 \gg \epsilon_q^2$) the total MSE is minimized if the prediction network performance is close to that of an optimized D*PCM coder and eq. (99) provides a lower bound in channel error variance for DPCM (and NFC) in the case of noisy channels. For low bit-error rates the best predictor will have a characteristic between full-whitening (error-free transmission) and half-whitening (D*PCM bound for noisy channels). Therefore

$$\min_{\text{PCM}} \{\sigma_c^2\} \geqslant \min_{\text{DPCM}} \{\sigma_c^2\} \geqslant \min_{\text{D*PCM}} \{\sigma_c^2\} \tag{102}$$

Example 6: Assume a previous-sample 2-bit DPCM and a Gaussian (not necessarily Markovian) source with adjacent-sample correlation $\rho = 0.85$. Let the bit-error probability be $P = 0.05$. From eq. (90) and Table IV we have $\epsilon_q^2 = 0.118$ and $\epsilon_c^2 = 0.233$. Figure 7 shows the dependence of the signal-to-noise ratio of this DPCM scheme on the value a of its predictor coefficient both for the noiseless and the noisy channel. The theoretical results have been calculated from eq. (101) using eqs. (16) and (31).

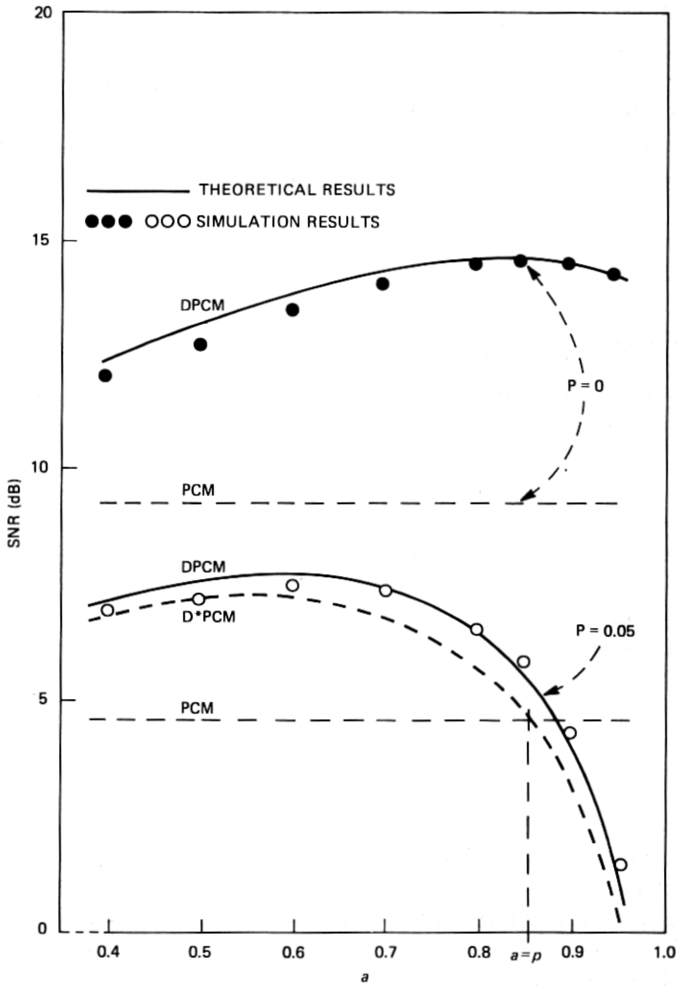

Fig. 7—DPCM performance on noisy channels. Two-bit quantization of Gaussian source with adjacent-sample correlation $\rho = 0.85$. Folded binary code with bit-error rates $P = 0$ and $P = 0.05$.

It is seen that DPCM performs better than PCM if the predictor is appropriately chosen. The differences between theory and measurements for low values of the predictor coefficient are again a consequence of the fact that the quantization error has been assumed to be not correlated with the input signal. For quantizers with a low number of levels this assumption only holds if the signal being quantized is uncorrelated, i.e. for predictor coefficients close to ρ. Notice that the crosspoint between PCM and DPCM performance is reached for a value of a which is slightly higher than ρ; this deviation from the predicted performance is a con-

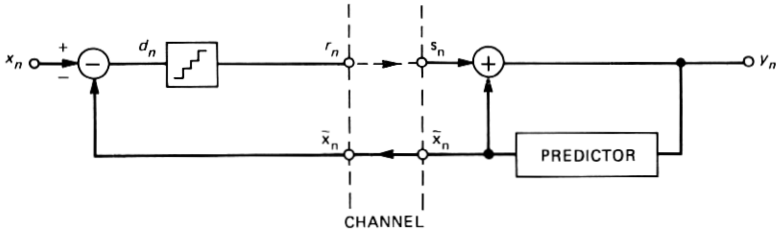

Fig. 8—Noiseless channel feedback DPCM.

sequence of the noise feedback. Figure 7 also compares the DPCM performance in the case of a noisy channel with the equivalent D*PCM performance which has been obtained from eqs. (98), (14), and (31). It is apparent that this one-tap D*PCM performance is a very useful bound of the DPCM performance. The optimal value of the DPCM predictor coefficient is also close to the optimal D*PCM coefficient a_{opt} as given in eq. (33) ($a_{opt} \approx 0.57$ for high bit-error probabilities). The choice of a_{opt} in accordance with eq. (33) for very noisy channels has first been mentioned in Ref. 41.

4.2.3 Noiseless channel feedback DPCM

In D*PCM the predictors of coder and decoder operate on slightly different signals, because there are quantizing noise and channel noise, respectively, in the intervening path. In DPCM both predictors operate on the same signal, viz. the sequence of reconstructed samples, if only the channel is noiseless. In the case of channel errors the predictions are again different, and the channel noise has the same effect on the overall MSE as in D*PCM. Let us assume that a noiseless channel from decoder to coder is available (see Fig. 8). It is then possible to ensure identical predictions via the feedback channel by retransmitting the prediction values calculated at the decoder. Notice that this scheme is identical to standard DPCM if the transmission of the encoded difference samples to the decoder is noiseless. In the case of channel errors, however, we have a total error

$$t_n = q_n + c_n \tag{103}$$

of variance

$$\sigma_t^2 = (\epsilon_q^2 + \epsilon_c^2) \cdot \sigma_d^2. \tag{104}$$

The contrast to D*PCM and standard DPCM is quite clear; the feedback is now around quantizer *plus* noisy channel, and thus error accumulation in the decoder loop has been totally avoided.

Example 7: In the case of previous-sample prediction the results of Example 1 (DPCM with analysis of the effects of noise feedback) apply. The optimum choice of the predictor-coefficient is given by eq. (17) if ϵ_q^2 is replaced with $\epsilon_q^2 + \epsilon_c^2$. To demonstrate the improvements let us use the figures of the previous example. We have $\epsilon_q^2 + \epsilon_c^2 = 0.351$, and hence $a_{opt} = 0.755$. We finally find [from eqs. (16) and (104)] a signal-to-noise ratio of 9.0 dB instead of 7.7 dB in the case of DPCM without noiseless channel feedback.

V. SUMMARY

There is a great interest in low bit-rate transmission of speech and television signals. Especially for acoustic signals it is well known that the subjective performance of a coder is strongly affected by the way in which quantizing distortion is distributed in frequency. In this paper we have compared coding schemes which employ prediction to exploit the inherent redundancy of these signals and which employ noise-shaping for optimizing the subjective performance on the basis of a frequency-weighted error criterion. First we have shown that DPCM outperforms D*PCM (a predictive scheme that lacks the feedback around the quantizer) for all nonwhite input spectra if the performance criterion is the unweighted total error variance. We have then used a noise-feedback coding structure as a framework for a unified analysis of predictive quantizing schemes. With this structure a minimum frequency-weighted error variance can be obtained by a proper shaping of the signal spectrum *and* the quantization noise prior to transmission. A comparison of this error variance with the distortion bound as given by the distortion-rate function for Gaussian signals has revealed that the performance of the noise-feedback coder is almost optimal for this class of signals. We have also shown that this coding structure degenerates to DPCM in the case of unweighted noise, and to D*PCM if the weighting function is inverse to the input spectrum. The performance results for these optimal coders have then been compared with those of suboptimal schemes including noise-feedback PCM and DPCM with prefiltering. For first-order Markov sources which often serve as a model of actual input spectra we have been able to derive simple explicit results in terms of the autocorrelation coefficient. In the last part we have examined the effects of channel transmission errors on the overall performance of these predictive quantizing schemes. We have shown that these coders when appropriately designed are less sensitive to channel errors than PCM. In D*PCM channel errors contribute to the total error in exactly the same way as quantization noise. Thus the D*PCM results provide a guideline for the optimization and a bound for the performance of DPCM.

VI. ACKNOWLEDGMENTS

The author wishes to thank A. Gersho, A. Wasiljeff, G. Wessels, and R. Zelinski for helpful suggestions. Thanks are also due to N. S. Jayant for significant contributions to the clarity of this paper.

REFERENCES

1. C. C. Cutler, "Transmission Systems Employing Quantization," U.S. Patent 2,927,962, 1960 (filed April 26, 1954).
2. J. L. Flanagan, *Speech Analysis, Synthesis, And Perception,* second edition, Berlin, Heidelberg, New York: Springer-Verlag, 1972.
3. J. V. Bodycomb and A. H. Haddad, "Some Properties of a Predictive Quantizing System," IEEE Trans. on Commun. Tech., *COM-18,* 1970, pp. 682–684.
4. J. P. Costas, "Coding with Linear Systems," Proc. IRE, *40,* 1952, pp. 1101–1103.
5. V. M. Shteyn, "On Design of Linear Predistortion and Correcting Systems," Radiotekn., *11,* No. 2, 1956, p. 60.
6. J. J. Spilker, "Theoretical Bounds on the Performance of Sampled Data Communications Systems," IRE Trans. Circuit Theory, *CT-7,* 1960, pp. 335–347.
7. R. A. Bruce, "Optimum Pre-Emphasis and De-Emphasis Networks for Transmission of Television by PCM," IEEE Trans. on Commun. Tech., *COM-12,* 1964, pp. 91–96.
8. L. M. Goodman and P. R. Drouilhet, "Asymptotically Optimum Pre-Emphasis and De-Emphasis Networks for Sampling and Quantizing," Proc. IEEE, *54,* 1966, pp. 795–796.
9. R. A. McDonald, Bell Laboratories, unpublished work.
10. B. G. Cramer, "Optimum Linear Filtering of Analog Signals in Noisy Channels," IEEE Trans. on Audio and Electroacoustics, *AU-14,* 1966, pp. 3–15.
11. T. Berger and D. W. Tufts, "Optimum Pulse Amplitude Modulation: Parts I and II," Trans. IEEE, *IT-13,* 1967, pp. 196–216.
12. T. Berger, *Rate Distortion Theory,* Englewood Cliffs, N.J.: Prentice Hall, Inc., 1971.
13. D. Chan and R. W. Donaldson, "Optimum Pre- and Postfiltering of Sampled Signals with Application to Pulse Modulation and Data Compression Systems," IEEE, Trans. on Commun. Tech., *COM-19,* No. 2, 1971, pp. 141–156.
14. E. G. Kimme and F. F. Kuo, "Synthesis of Optimal Filters for a Feedback Quantization System," IEEE Trans. on Circuit Theory, *CT-10,* 1963, pp. 405–413.
15. R. C. Brainard and J. C. Candy, "Direct-Feedback Coders: Design and Performance with Television Signals," Proc. IEEE, *57,* No. 5, 1969, pp. 776–786.
16. K. Nitadori, "Statistical Analysis of Δ-PCM," J. Inst. Electron. Commun. Eng., Japan, *48,* 1965, pp. 17–26.
17. R. A. McDonald, "Signal-to-Noise and Idle Channel Performance of Differential Pulse Code Modulation Systems—Particular Applications to Voice Signals," B.S.T.J., *45,* No. 7 (September 1966), pp. 1123–1151.
18. P. Noll, "Nonadaptive and Adaptive Differential Pulse Code Modulation of Speech Signals," Polytechnisch Tijdschrift, Den Haag 1972, No. 19, pp. 623–629.
19. P. Noll, Bell Laboratories, unpublished work.
20. U. Grenander and G. Szegö, "Toeplitz Forms and their Applications," Berkeley, Calif.: University of California Press, 1958.
21. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech,* New York: Springer-Verlag, 1976.
22. P. Noll, "Some Properties of Differential PCM Schemes" (in German), Archiv für Elektronik und Übertragungstechnik (AEÜ), Bd. 30, 1976, pp. 125–130.
23. P. Noll, "Transform Coding of Speech," International Conference on Communications 1977, Conference Record Vol. 1, pp. 13.5-306–13.5-309.
24. D. J. Esteban and J. E. Menez, "Low Bit Rate Voice Transmission Based on Transversal Block Coding," Acoustical Society of America, 91st ASA Meeting, 1976, Washington, Paper RR 15.
25. H. G. Musmann, "Redundancy Reduction by Linear Transforms" (in German), Nachrichtentechnischer Fachbericht, NTF Band 40, 1970, pp. 13–27.
26. K. W. Cattermole, *Principles of Pulse Code Modulation,* Iliffe Books, Ltd., 1969.
27. W. Schlink, "On Source Encoding of PCM Speech Signals" (in German), Thesis, TU Braunschweig, 1976.

28. J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, *63,* 1975, pp. 561–580.
29. H. A. Spang and P. M. Schultheiss, "Reduction of Quantizing Noise by Use of Feedback," IRE Trans. on Commun. Systems, 1962, pp. 373–380.
30. B. Heuser and P. Noll, "Error Feedback as a Means For Colouring Quantization Noise" (in German), Heinrich-Hertz-Institut Berlin, Tech. Rep. No. 191, 1976.
31. J. B. O'Neal, "Bounds on Subjective Performance Measures for Source Encoding Systems," IEEE Trans. on Inform. Theory, *IT-17,* 1971, pp. 224–231.
32. J. M. Barstow and H. N. Christopher, "Measurement of Random Video Interference to Monochrome and Color TV," Trans. AIEE, Commun. and Electronics, 1962, pp. 313–320.
33. R. L. Dobrushin and B. S. Tsybakov, "Information Transmission with Additional Noise," IRE Trans. Inform. Theory, *IT-8,* 1962, pp. 293–304.
34. R. A. McDonald and P. M. Schultheiss," Information Rates of Gaussian Signals under Criteria Constraining the Error Spectrum," Proc. IEEE (Correspondence), *52,* 1964, pp. 415–416.
35. R. E. Totty and G. C. Clark, "Reconstruction Error in Waveform Transmission," IEEE Trans. Inform. Theory (Correspondence), *IT-13,* 1967, pp. 336–338.
36. J. Max, "Quantizing for Minimum Distortion," IRE Trans. on Inform. Theory, *IT-6,* 1960.
37. A. J. Kurtenbach and P. A. Wintz, "Quantizing for Noisy Channels," IEEE Trans. on Communications, *COM-17,* No. 2 (1969), pp. 291–302.
38. B. R. Murthy, "Optimization in PCM, Adaptive PCM and DPCM Systems," Purdue University, Thesis, 1969.
39. P. Noll, "Effects of Channel Errors on the Signal-to-Noise Performance of Speech-Encoding Systems," B.S.T.J., 54, No. 9 (November 1975), pp. 1615–1637.
40. A. J. Viterbi, "Lower Bounds on Maximum Signal-to-Noise Ratios for Digital Communication Over the Gaussian Channel," IEEE Trans. on Communications, *COM-12* (1964), pp. 10–17.
41. K. Chang and R. W. Donaldson, "Analysis, Optimization and Sensitivity Study of Differential PCM Systems Operating on Noisy Communication Channels," IEEE Trans. on Communications, *COM-20,* No. 3 (1972), pp. 338–350.
42. J. E. Essman and P. A. Wintz, "The Effects of Channel Errors in DPCM Systems and Comparison with PCM Systems," IEEE Trans. on Communications, *COM-21,* No. 8 (1973), pp. 867–877.