Achieving the Gaussian Rate-Distortion Function by Prediction

Ram Zamir, Yuval Kochman and Uri Erez
Dept. Electrical Engineering-Systems, Tel Aviv University

***Abstract*— The "water-filling" solution for the quadratic rate-distortion function of a stationary Gaussian source is given in terms of its power spectrum. This formula naturally lends itself to a frequency domain "test-channel" realization. We provide an alternative time-domain realization for the rate-distortion function, based on linear prediction. This solution has some interesting implications, including the optimality at all distortion levels of vector-quantized differential pulse code modulation (DPCM), and a duality relationship with decision-feedback equalization (DFE) for inter-symbol interference (ISI) channels.**

I. Introduction: Rate-Distortion and Prediction

The rate-distortion function (RDF) of a stationary source with memory is given by a *time domain* formula, that is, as a limit of normalized mutual informations associated with vectors of source samples. For a real valued source $\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$, and mean-squared distortion level D, the RDF can be written as, [1],

$$\overline{R}(D) = \lim_{n\to\infty} \frac{1}{n} \inf I(X_1, \ldots, X_n; Y_1, \ldots, Y_n)$$

where the infimum is over all channels $\mathbf{X} \to \mathbf{Y}$ such that $\frac{1}{n}\|\mathbf{Y} - \mathbf{X}\|^2 \le D$. A channel which realizes this infimum is called an *optimum test-channel*.

When the source is Gaussian, the RDF takes an explicit form in the *frequency domain*. Assuming auto-correlation function $R[k] = E\{X_n X_{n+k}\}$, and power-spectrum

$$S(e^{j2\pi f}) = \sum_k R[k] e^{-jk2\pi f}, \quad -\pi < f < \pi$$

the RDF is given by the "water filling" formula

$$\begin{aligned} \overline{R}(D) &= \int_{-1/2}^{1/2} \frac{1}{2} \log\left(\frac{S(e^{j2\pi f})}{D(e^{j2\pi f})}\right) df \\ &= \int_{f:S(e^{j2\pi f})>\theta} \frac{1}{2} \log\left(\frac{S(e^{j2\pi f})}{\theta}\right) df \end{aligned} \quad (1)$$

where $D(e^{j2\pi f})$ is the distortion spectrum

$$D(e^{j2\pi f}) = \begin{cases} \theta, & \text{if } S(e^{j2\pi f}) > \theta \\ S(e^{j2\pi f}), & \text{otherwise,} \end{cases} \quad (2)$$

and $\theta = \theta(D)$ is the "water level" chosen such that $\int_{-1/2}^{1/2} D(e^{j2\pi f}) df = D$. For the special case of a memoryless (white) Gaussian source $\sim N(0, \sigma^2)$, the power-spectrum is flat $S(e^{j2\pi f}) = \sigma^2$, and the RDF is simplified to

$$\frac{1}{2} \log\left(\frac{\sigma^2}{D}\right). \quad (3)$$

Our main result is a predictive "time-domain" realization for the quadratic-Gaussian RDF (1). The notion of *entropy-power* and the *Shannon lower bound* provide a simple relation between (1) and prediction, which motivates our result. Recall that that the entropy-power is the variance of a *white* Gaussian process having the same entropy-rate as the source; for a Gaussian source the entropy-power is given by

$$P_e(X) = \exp\left(\int_{-1/2}^{1/2} \log\left(S(e^{j2\pi f})\right) df\right). \quad (4)$$

The Shannon lower bound states that for any D

$$\overline{R}(D) \ge \frac{1}{2} \log\left(\frac{P_e(X)}{D}\right) \quad (5)$$

with equality for distortion levels smaller than or equal to the lowest value of the power spectrum: $D \le \min_f S(e^{j2\pi f})$ (in which case $D(e^{j2\pi f}) = \theta = D$). See [1]. In the context of Wiener's spectral-factorization theory, (4) quantifies the mean-squared error of the one-step linear predictor of the source from its infinite past [1]; that is, we also have

$$P_e(X) = \inf_{\{a_i\}} E\left(X_n - \sum_{i=1}^{\infty} a_i X_{n-i}\right)^2. \quad (6)$$

Now, by the orthogonality principle (see [4]), the error process of the optimum infinite-order predictor (sometimes called the "innovation process")

$$E_n = X_n - \sum_{i=1}^{\infty} a_i X_{n-i}$$

has zero mean and is *white*. Hence, in view of (3) and (5), for small distortion levels the RDF of a Gaussian source with memory is equal to the RDF of its *memoryless* prediction error process at the same distortion level.

We shall see later in Section II how the observation above translates into a predictive test-channel which can realize the RDF not only for small but for *all* distortion levels. Before that, we'd like to consider another feature of the frequency-domain solution (1) which motivates the predictive time-domain result of Section II. Note that the optimum test-channel that realizes the memoryless RDF (3) takes the form of a memoryless linear-additive noise channel:

$$Y = \beta(\alpha X + N)$$

with $\alpha = \beta = \sqrt{1 - D/\sigma^2}$ and $N \sim N(0, D)$. (In the general case α and β take the form of pre- and post-filters.) In view

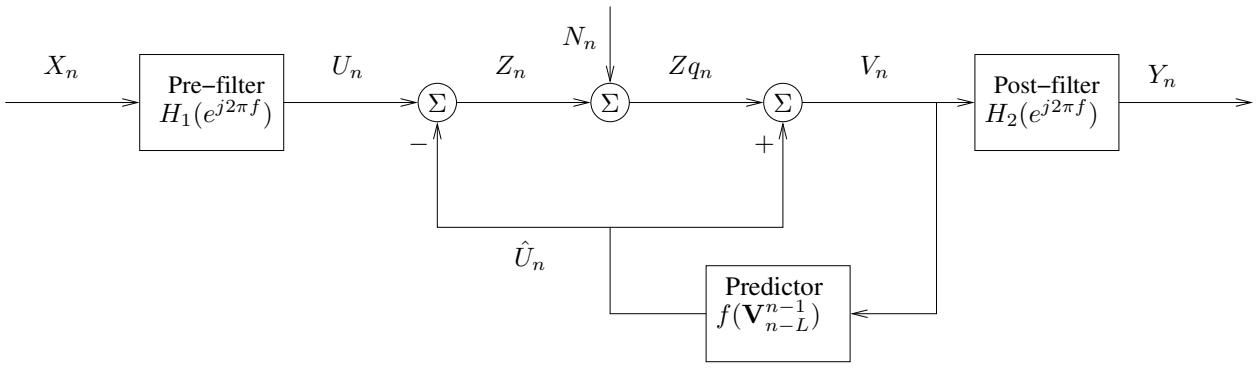

Fig. 1. Predictive Test Channel

of that, one way to understand the form of the general RDF in (1) is to look on its discrete approximation

$$\sum_i \frac{1}{2}\log\left(\frac{S(e^{j2\pi f_i})}{D(e^{j2\pi f_i})}\right) \tag{7}$$

as encoding of *parallel* (independent) Gaussian sources, where source i is a memoryless Gaussian source $X_{f_i} \sim N(0, S(e^{j2\pi f_i}))$ encoded at distortion $D(e^{j2\pi f_i})$; see [3]. Furthermore, the RDF (1) can be realized by a vector of parallel channels of the form $Y_{f_i} = \beta_i(\alpha_i X_{f_i} + N_i)$. This interpretation motivates practical, frequency domain source coding schemes such as Transform Coding and Sub-band Coding [5], which get close to the RDF of a Gaussian source with memory using *scalar* quantization.

Section II proposes an alternative formulation for the RDF (1) that is motivated by the time-domain quantization scheme of Differential Pulse Code Modulation (DPCM). The goal of the new formulation is, like in the frequency domain interpretation of (7), to translate the encoding of dependent source samples into a series of encodings of *independent* sources. The task of removing the dependence in the time domain approach is achieved by prediction.

Kim and Berger [8] showed that we cannot achieve the RDF of an auto-regressive (AR) Gaussian process by encoding its innovation process. Their scheme amounts to *open-loop prediction* of the source. In this paper we show that RDF can be achieved by embedding the encoder inside the prediction loop, i.e., by *closed-loop prediction*.

After presenting and proving our main result in Sections II and III, respectively, we provide reflections on the result and its operational implications. Section IV discusses the spectral features of the solution, Section V relates it to compression of parallel sources, and Section VI discusses implementation by Entropy Coded Dithered Quantization (ECDQ). Finally, in Section VII we relate prediction in source coding to prediction in the area of channel equalization and to recent observations by Forney [4]. Like in [4], our analysis is done mainly using the properties of information measures; from Wiener estimation theory we need only two basic results: the orthogonality principle and the one-step prediction error formula (6).

II. Main Result: A Predictive Test Channel

Consider the system in Figure 1, which consists of three basic blocks: pre-filter $H_1(e^{j2\pi f})$, a noisy channel embedded in a close loop, and a post-filter $H_2(e^{j2\pi f})$. The system parameters are derived from the water-filling solution (1)-(2). The source samples $\{X_n\}$ are passed through a pre-filter, whose phase is arbitrary and its absolute squared frequency response is given by

$$|H_1(e^{j2\pi f})|^2 \;=\; 1 - \frac{D(e^{j2\pi f})}{S(e^{j2\pi f})}. \tag{8}$$

The pre-filter output, denoted U_n, is being fed to the central block which generates a process V_n according to the the following recursion equations:

$$\hat{U}_n \;=\; f(V_{n-1}, V_{n-2}, \ldots, V_{n-L}) \tag{9}$$
$$Z_n \;=\; U_n - \hat{U}_n \tag{10}$$
$$Zq_n \;=\; Z_n + N_n \tag{11}$$
$$V_n \;=\; \hat{U}_n + Zq_n \tag{12}$$

where $N_n \sim N(0,\theta)$ is zero-mean additive white Gaussian noise (AWGN) independent of the input process $\{U_n\}$ whose variance is equal to the "water level" θ, and $f(\cdot)$ is some "prediction function" for the input U_n given the L "past" samples of the output process $(V_{n-1}, V_{n-2}, \ldots, V_{n-L})$. Finally, the post-filter frequency response is the complex conjugate of that of the pre-filter,

$$H_2(e^{j2\pi f}) = H_1^*(e^{j2\pi f}). \tag{13}$$

The block from U_n to V_n is equivalent to the configuration of differential pulse code modulation (DPCM), [7], [5], with the DPCM (scalar) quantizer replaced by the AWGN channel $Zq_n = Z_n + N_n$. In particular, by combining the recursion equations (9)-(12) it follows that this block satisfies the well known "DPCM error identity", [7],

$$V_n = U_n + (Zq_n - Z_n) = U_n + N_n. \tag{14}$$

That is, the output V_n is a noisy version of the input U_n via the AWGN channel $V_n = U_n + N_n$. In DPCM the prediction function f is linear:

$$f(V_{n-1}, \ldots, V_{n-L}) = \sum_{i=1}^{L} a_i V_{n-i} \tag{15}$$

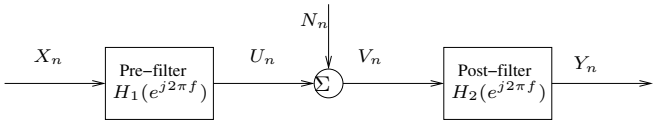

Fig. 2. Equivalent Channel

where $a_1, \ldots, a_L$ minimize the mean-squared "noise" prediction error

$$E\left(U_n - \sum_{i=1}^{L} a_i V_{n-i}\right)^2. \tag{16}$$

If $\{U_n\}$ and $\{V_n\}$ are jointly Gaussian, then the best predictor of any order is linear, so the minimum of (16) is also the the minimum mean squared error (MMSE) in estimating U_n from the vector $(V_{n-1}, \ldots, V_{n-L})$. We shall further elaborate on the relationship with DPCM later.

Note that while the central block is sequential and hence causal, the pre- and post-filters are non-causal and therefore their realization in practice requires large delay. Our main result is the following.

Theorem 1: For any stationary source with power spectrum $S(e^{j2\pi f})$, the system of Figure 1 satisfies

$$E(Y_n - X_n)^2 = D. \tag{17}$$

Furthermore, if the source X_n is Gaussian, and if the function $f(\cdot)$ is the optimum linear predictor of U_n given the *infinite* past $(V_{n-1}, V_{n-2}, \ldots)$, i.e., $L = \infty$ in (9)-(12), then for all n

$$I(Z_n; Z_n + N_n) = \overline{R}(D). \tag{18}$$

The proof is given in Section III. The main feature of Theorem 1 is the fact that the left-hand side of (18) is a *single* letter mutual information. Thus, in a sense the core of the encoding process amounts to a *memoryless* AWGN test-channel.

Another interesting feature of the system is the relationship between the prediction error process Z_n and the original process X_n. If X_n is an auto-regressive (AR) process, then in the limit of small distortion ($D \to 0$), Z_n is roughly its innovation process. Hence, unlike in open-loop prediction [8], encoding the innovations in a closed-loop system is optimal in the limit of high-resolution encoding. We shall return to that, as well as discuss the case of general resolution, in Section IV.

III. Proof of Main Result

Note first that the error identity (14) implies that the entire system of Figure 1 is equivalent to the system depicted in Figure 2, consisting of a pre-filter (8), an AWGN channel with noise variance θ, and a post-filter (13). This is, in fact, one of the equivalent forms of the *forward channel realization* of the quadratic-Gaussian RDF [1, Section 4.5], [9]. In particular, simple spectral analysis shows that the power spectrum of the overall error process $Y_n - X_n$ is equal to the water filling distortion spectrum $D(e^{j2\pi f})$ in (2); hence the total distortion in (17) is D as desired.

For the second part, since the system of Figure 2 coincides with the forward channel realization of the quadratic-Gaussian RDF, for a Gaussian source we have

$$\overline{I}(\{X_n\}; \{Y_n\}) = \overline{I}(\{U_n\}; \{V_n\}) = \overline{R}(D)$$

where $\overline{I}$ denotes mutual information-rate between jointly stationary sources:

$$\overline{I}(\{X_n\}; \{Y_n\}) = \lim_{n\to\infty} \frac{1}{n} I(X_1, \ldots, X_n; Y_1, \ldots, Y_n). \tag{19}$$

Hence, the proof will be completed by showing that

$$\overline{I}(\{U_n\}; \{V_n\}) = I(Z_n; Z_n + N_n). \tag{20}$$

To that end, consider the conditional mutual information between $U_1, \ldots, U_n$ and $V_1, \ldots, V_n$ given $V_{-L}^0 = V_0, V_{-1}, \ldots, V_{-L}$. Using the chain rule we have

$$\begin{aligned} &I(U_1, \ldots, U_n; V_1, \ldots, V_n | V_{-L}^0) \\ &\qquad = \sum_{i=1}^{n} I(U_1, \ldots, U_n; V_i | V_{-L}^{i-1}). \end{aligned}$$

By the recursion equations (9)-(12), the i-th term in the sum can be written as

$$\begin{aligned} & I(U_1, \ldots, U_n; V_i | V_{-L}^{i-1}) \\ = \;& I(U_1, \ldots, U_{i-1}, U_i - \hat{U}_i, U_{i+1}, \ldots, U_n; V_i - \hat{U}_i | V_{-L}^{i-1}) \\ = \;& I(U_1, \ldots, U_{i-1}, Z_i, U_{i+1}, \ldots, U_n; Z_i + N_i | V_{-L}^{i-1}). \end{aligned}$$

Since the AWGN N_n is independent of the source $U_1, \ldots, U_n$ and of the past values of V_n, we have the Markov chain relation

$$Z_i + N_i \Leftrightarrow (Z_i, V_{-L}^{i-1}) \Leftrightarrow (U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_n).$$

(Note that future values of the output process $V_{i+1}, V_{i+2}, ..$ depend on N_i, but this does not affect the Markov chain above.) This relation implies that the i-th term above simplifies to

$$I(Z_i; Z_i + N_i | V_{-L}^{i-1}).$$

Until now we didn't need to use the source Gaussianity nor the optimality of the prediction function $f(\cdot)$. Taking these into account, and letting the predictor order $L \to \infty$, the *orthogonality principle* of MMSE estimation implies that the *estimation error* Z_i is statistically independent of the *measurements* $(V_{i-1}, V_{i-2}, \ldots)$. Since N_i is independent of both Z_i and the past V_i's, the i-th term above further simplifies to $I(Z_i; Z_i + N_i)$. We obtain

$$\begin{aligned} &I(U_1, \ldots, U_n; V_1, \ldots, V_n | V_0, V_{-1}, \ldots) \\ &= \sum_{i=1}^{n} I(Z_i; Z_i + N_i) = n I(Z_n; Z_n + N_n) \end{aligned}$$

where the second equality follows since the system is time-invariant and all the processes are stationary. Using the definition of mutual information rate (19), and noting that it remains unchanged if we condition on the infinite past, we arrive at (20) and the proof is completed.

IV. Properties of the Predictive Test-Channel

The following observations shed light on the behavior of the test channel of Figure 1.

Prediction in the high resolution regime. If the power-spectrum $S(e^{j2\pi f})$ is everywhere positive (e.g., if $\{X_n\}$ can be represented as an AR process), then in the limit of small distortion $D \to 0$, predicting U_n from its "noisy past" $\{V_i = U_i + N_i : \ i = n-1, n-2, \ldots\}$, is equivalent to predicting U_n from its "clean past" $\{U_{n-1}, U_{n-2}, \ldots\}$. Hence, in this limit the prediction error Z_n is equal to the "innovation process" associated with U_n. Since for small distortion the pre- and post-filters (8), 13) are roughly all-pass filters, U_n has roughly the same power spectrum as X_n; hence Z_n is in fact equivalent to the innovation process of X_n. In particular, Z_n is an i.i.d. process whose variance is $P_e(X)$ = the entropy-power of the source (4).

Prediction in the general case. Interestingly, for general distortion $D > 0$, the prediction error Z_n is *not white*, as the noisiness of the past does not allow the predictor f to remove all the source memory. Nevertheless, the noisy version of the prediction error $Zq_n = Z_n + N_n$ is white for every $D > 0$, because it amounts to predicting V_n from its own infinite past: since N_n is white and has zero-mean, $\hat{U}_n$, which is the optimal predictor for U_n, is also the optimal predictor for $V_n = U_n + N_n$. This whiteness might seem at first a contradiction, because Zq_n is the sum of a non-white process Z_n and a white process N_n. However, $\{Z_n\}$ and $\{N_n\}$ are *not* independent processes, because Z_n depends on past values of N_n (through the past of V_n). Hence the channel $Zq_n = Z_n + N_n$ is not quite an additive-noise channel; rather, it is *sequentially-additive*: the noise is independent of past and present channel inputs, but is not independent of future inputs. These observations imply that the channel $Zq_n = Z_n + N_n$ satisfies:

$$I(Z_n; Z_n + N_n | Z_1 + N_1, ..., Z_{n-1} + N_{n-1}) = I(Z_n; Z_n + N_n) \ ,$$

while in general

$$\bar{I}(\{Z_n\}; \{Z_n + N_n\}) > I(Z_n; Z_n + N_n) \ .$$

The channel when the Shannon lower bound holds. As long as D is smaller than the lowest point of the source spectrum (i.e., $D(e^{j2\pi f}) = \theta = D$ in (1)), the quadratic-Gaussian RDF coincides with the Shannon Lower Bound (5). In this case, the following properties hold for the predictive test channel:

- The power spectra of U_n and Y_n are the same and are equal to $S(e^{j2\pi f}) - D$.
- The power spectrum of V_n is equal to the power spectrum of the source $S(e^{j2\pi f})$.
- Since the (white) process $Z_n + N_n$ is the optimal prediction error of the process V_n from its *own* infinite past, its variance is $P_e(V) = P_e(X)$ of (4).
- As a consequence we have

$$I(Z_n; Z_n + N_n) = h(N(0, P_e(X))) - h(N(0, D)))$$
$$= 1/2 \log(P_e(X)/D)$$

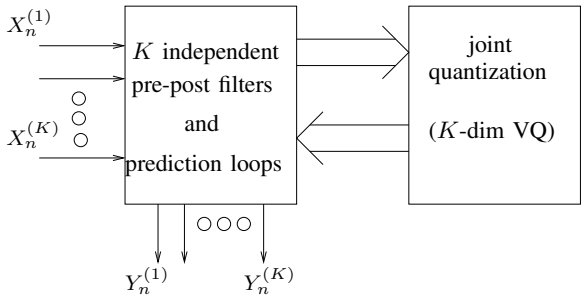

Fig. 3. Parallel sources.

which is indeed the Shannon lower bound (5).

V. Vector-Quantized DPCM and D*PCM

As mentioned, the structure of the loop in the channel of Figure 1 is of a DPCM quantizer, with the scalar quantizer replaced by the additive noise. However, if we wish to implement the additive noise by a quantizer realizing the full *single letter* mutual information $I(Z_n; Z_n + N_n)$, we must use *vector* quantization (VQ). It is not possible to do that along the time domain, due to the sequential nature of the system above. Nevertheless, we can achieve the VQ gain by adding a "spatial" dimension, if we jointly encode a large number of parallel sources, as happens, e.g., in video coding. See Figure 3.

If we have only one source with decaying memory, we can still achieve the rate distortion function at the cost of large delay, by using interleaving.

If we do not use any of the above, but restrict ourselves to scalar quantization, we have a pre/post filtered DPCM scheme. It follows from Theorem 1 that in principle, a pre/post filtered DPCM scheme is optimal (up to the loss of the VQ gain) at all distortion levels, and not only at high resolution. It is interesting to mention that In the quantization literature, the "open loop" prediction approach investigated in [8] is referred to as D*PCM [7].

VI. A Dual Relationship with Decision-Feedback Equalization

Consider the (real-valued) discrete-time time-invariant linear Gaussian channel arising at the output of a sampled matched filter,

$$Y_n = \sum_{k=-\infty}^{\infty} h_k X_{n-k} + Z_n. \tag{21}$$

Here h_n is the equivalent discrete-time impulse response resulting from the cascade of the spectral shaping filter, the modulation pulse, the continuous-time channel as well as the matched filter. It follows that $H(e^{j2\pi f})$ is non-negative and the autocorrelation function of Z_n is

$$R_{ZZ}(k) = E\{Z_{n+k} Z_n\} = \frac{N_0}{2} h_k. \tag{22}$$

Let X_n be an iid Gaussian random process with power σ_x^2. Then the mutual information (normalized per symbol) between

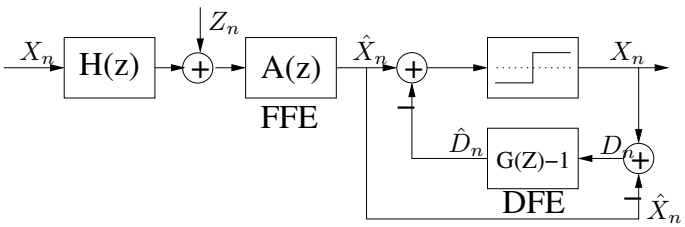

Fig. 4. MMSE-DFE in predictive form.

the input and output of the channel is

$$\overline{I}(\{X_n\},\{Y_n\}) = \int_{-1/2}^{1/2} \frac{1}{2}\log\left(1+\frac{\sigma_x^2 H(e^{j2\pi f})}{N_0/2}\right) df. \quad (23)$$

Capacity is achieved by using a spectral shaping filter satisfying the "water-filling" power allocation. As reflected in the expression (23), capacity may be achieved by parallel AWGN coding over narrow frequency bands as done in practice in DMT/OFDM systems. A well known alternative capacity-achieving scheme which is based on prediction rather than the Fourier transform is offered by the canonical MMSE-DFE equalization structure used in single-carrier transmission. These observations parallel those made in Section I with respect to the RDF.

It is well known that the capacity of linear Gaussian channels can be achieved using MMSE-DFE coupled with AWGN coding. This has been shown using different approaches. One such approach which is particularly illuminating in our context is based on linear prediction of the error sequence. We now recount this result which was developed and refined by numerous authors, see [6], [2], [4] and references therein. Our exposition closely follows that of Forney [4].

As a first step, let $\hat{X}_n$ be the optimal MMSE estimator of X_n from the channel output sequence $\{Y_n\}$. Since $\{X_n\}$ and $\{Y_n\}$ are jointly Gaussian and stationary this estimator is linear and time invariant. Denote the estimation error, which is composed of residual ISI and Gaussian noise, by D_n. Then

$$X_n = \hat{X}_n + D_n \quad (24)$$

where $\{D_n\}$ is independent of $\{\hat{X}_n\}$ due to the orthogonality principle and Gaussianity.

Assuming correct decoding of past symbols[1], the decoder knows the past samples $D_{n-1}, D_{n-2}, \ldots$ and may form an optimal predictor, $\hat{D}_n$, of the estimation error D_n. The prediction error $E_n = D_n - \hat{D}_n$ has variance $P_e(D)$, the entropy power of D_n. This prediction may then be added to $\hat{X}_n$ to form $\tilde{X}_n$. It follows that

$$\begin{aligned} X_n &= \hat{X}_n + D_n \\ &= \tilde{X}_n - \hat{D}_n + D_n \\ &= \tilde{X}_n + E_n, \end{aligned} \quad (25)$$

[1]Here we must actually break with assumption that X_n is a Gaussian process. and We implicitly assume that X_n are symbols of a capacity-achieving AWGN code. The slicer should be viewed as a mnemonic aid where in practice an optimal decoder should be used. Furthermore, the use of an interleaver and long delay is necessary. See [4]

where $\{\tilde{X}_n\}$ and $\{E_n\}$ are statistically independent. It follows that the residual estimation error satisfies

$$E\{X_n - \tilde{X}_n\}^2 = E\{D_n - \hat{D}_n\}^2 = P_e(D), \quad (26)$$

The channel (25) is often referred to as the *backward channel*. Furthermore, since X_n and E_n are i.i.d Gaussian, it is an AWGN channel. We have therefore derived the following.

Theorem 2: The mutual information of the channel (21) is equal to the scalar mutual information

$$I(\tilde{X}_n; \tilde{X}_n + E_n)$$

of the memoryless channel (25).

Proof: Let $X_n^- = \{X_{n-1}, X_{n-2}, \ldots\}$ and $D_n^- = \{D_{n-1}, D_{n-2}, \ldots\}$. Using the chain rule of mutual information we have

$$\begin{aligned} \overline{I}(\{X_n\},\{Y_n\}) &= h(\{X_n\}) - h(X_n|\{Y_n\}, X_n^-) \\ &= h(\{X_n\}) - h(X_n - \hat{X}_n|\{Y_n\}, X_n^-) \\ &= h(\{X_n\}) - h(D_n|\{Y_n\}, X_n^-) \\ &= h(\{X_n\}) - h(D_n|\{Y_n\}, D_n^-, X_n^-) \\ &= h(\{X_n\}) - h(D_n - \hat{D}_n|\{Y_n\}, D_n^-) \\ &= h(\{X_n\}) - h(E_n|\{Y_n\}, D_n^-) \\ &= h(\{X_n\}) - h(E_n) \quad (27) \\ &= I(\tilde{X}_n; \tilde{X}_n + E_n), \end{aligned}$$

where (27) follows from successive application of the orthogonality principle [4]. ■

It follows that

$$\int_{-1/2}^{1/2} \frac{1}{2}\log\left(1+\frac{\sigma_x^2 H(e^{j2\pi f})}{N_0/2}\right) df = \frac{1}{2}\log\left(\frac{\sigma_x^2}{P_e(D)}\right). \quad (28)$$

As a corollary, from (23), Theorem 2 and (28), we obtain the following well known result from Wiener theory,

$$P_e(D) = \exp\left(\int_{-1/2}^{1/2} \log\left(\frac{\sigma_x^2 \frac{N_0}{2}}{\sigma_x^2 H(e^{j2\pi f}) + \frac{N_0}{2}}\right) df\right).$$

REFERENCES

[1] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[2] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. J. Forney. MMSE Decision-Feedback Equalizers and Coding - Part I: Equalization Results. *IEEE Trans. Communications*, COM-43:2582–2594, Oct. 1995.

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

[4] G. D. Forney, Jr. Shannon meets Wiener II: On MMSE estimation in successive decoding schemes. In *42st Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois*, Oct. 2004.

[5] G.D.Gibson, T.Berger, T.Lookabaugh, D.Lindbergh, and R.L.Baker. *Digital Compression for Multimedia: Principles and Standards*'. Morgan Kaufmann Pub., San Fansisco, 1998.

[6] T. Guess and M. K. Varanasi. An information-theoretic framework for deriving canonical decision-feedback receivers in gaussian channels. *IEEE Trans. Information Theory*, IT-51:173–187, Jan. 2005.

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

[8] K. T. Kim and T. Berger. Sending a Lossy Version of the Innovations Process is Suboptimal in QG Rate-Distortion. In *Proceedings of ISIT-2005, Adelaide, Australia*, pages 209–213, 2005.

[9] R. Zamir and M. Feder. Information rates of pre/post filtered dithered quantizers. *IEEE Trans. Information Theory*, pages 1340–1353, September 1996.