

# Prediction-Based Lower Triangular Transform

See-May Phoong, *Member, IEEE*, and Yuan-Pei Lin, *Member, IEEE*

**Abstract**—In this paper, a new nonunitary transform called the prediction-based lower triangular transform (PLT) is introduced for signal compression. The new transform has the same decorrelation property as the Karhunen–Loeve transform (KLT), but its implementational cost is less than one half of KLT. Compared with the KLT, the design cost of an  $M \times M$  PLT is much lower and is only of the order of  $\mathcal{O}(M^2)$ . Moreover, the PLT can be factorized into simple building blocks. Using two different factorizations, we introduce two minimum noise structures that have roughly the same complexity as the direct implementation of PLT. These minimum noise structures have the following properties: 1) Its noise gain is unity even though the transform is nonunitary; 2) perfect reconstruction is structurally guaranteed; 3) it can be used for both lossy/lossless compression. We will show that the coding gain of PLT implemented using the minimum noise structure is the same as that of KLT. Furthermore, universal transform coders using PLT are derived. For AR(1) process, the  $M \times M$  PLT has a closed form and needs only  $(M - 1)$  multiplications and additions.

**Index Terms**—Compression, Karhunen–Loeve transform (KLT), subband coding, transform coding, wavelet coding.

## I. INTRODUCTION

TRANSFORM coding has played an important role in various areas of signal processing and communication [1]–[3]. It has been widely applied to data compression. It is well known that given the input statistics, the optimal unitary transform that yields the maximum coding gain is the Karhunen–Loeve transform (KLT) or Hotelling transform. The KLT is a unitary matrix that consists of the eigenvectors of input autocorrelation matrix. Due to its signal dependence and computational cost, the KLT is often only used as a benchmark for performance comparison. In many applications, suboptimal but signal independent transforms like the discrete cosine transform (DCT) are often used.

Recently, universal transform coding schemes using KLT have been developed. In [4], the authors consider the problem of universal transform coding based on estimating the KLT from quantized data. No side information is needed because both encoder and decoder can access the quantized data. Promising experimental results are demonstrated. In [5], the authors introduce a classification-based method using the KLT. The signal space is divided into a number of classes, and a fixed transform is designed for each class. In the proposed two-stage algorithm,

the encoder uses a collection of transform/bit allocation pairs. Very good coding performance is demonstrated.

The transform coders can be considered to be a subclass of orthonormal subband coders, which have the advantage of having unity noise gain. The orthonormal subband coders have coding gain  $\geq 1$  for any input statistics. Recently, there has been great interest in designing orthonormal subband coders that maximize the coding gain for given input statistics [6]–[10]. The problem of ideal optimal orthonormal coder has recently been solved. It is shown [6] that the optimal orthonormal filter bank is closely related to the principle component filter bank [7], [10]. An ideal orthonormal coder is optimal if and only if it satisfies the majorization and decorrelation properties [6]. The FIR case is studied in [8] and [9].

In many applications, it is desired that a lossy coding system becomes lossless when a sufficient bit rate is available. Several lossy/lossless coding systems have been proposed recently. In [11] and [12], the ladder structure is applied to high-fidelity compression of medical images. In [13], the authors introduce a new transform called the S+P transform. It is demonstrated that in the application of both lossy and lossless image coding, the S+P transform produces excellent compression results. In [14], the optimal predictor with certain zero constraint is used, and the filter is obtained through the optimization of Bernstein polynomial. In [15], the authors propose an integer-to-integer transform based on the ladder structure for lossless coding of images. Image coding using a two-dimensional (2-D) four-channel ladder structure is studied in [16]. However, like most biorthogonal coders, none of these coders has the unity noise gain property. Therefore, in the case of lossy compression, the coding gain of these proposed coders is not guaranteed to be greater than unity. In [17], we introduce a minimum noise structure for two-channel ladder-based filter banks. The minimum noise structure ensures that the noise gain is unity, even though the filter bank is never orthonormal. The coding gain can never be smaller than unity.

In this paper, a new transform called the prediction-based lower triangular transform (PLT) is introduced. The PLT is a signal-dependent nonunitary transform. It has the same decorrelation property as the KLT, and its coding gain is the same as the KLT. The elements of the PLT matrix are the coefficients of prediction polynomials of different orders. In addition to its coding performances, the PLT has many other merits.

- 1) The implementational cost is less than one half of the KLT.
- 2) The design cost of an  $M \times M$  PLT is only in the order of  $\mathcal{O}(M^2)$ .
- 3) It has a structurally perfect reconstruction (PR) implementation. Furthermore, PR is preserved even when all the multipliers are quantized.

Manuscript received November 24, 1998; revised December 1, 1999. This work was supported by the National Science Council under Grants NSC 89-2213-E-002-102 and NSC 89-2213-E-009-118, Taiwan, R.O.C. The associate editor coordinating the review of this paper and approving it for publication was Dr. Hitoshi Kiya.

S.-M. Phoong is with the Department of Electrical Engineering, Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (e-mail: smp@cc.ee.ntu.edu.tw).

Y.-P. Lin is with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

Publisher Item Identifier S 1053-587X(00)04922-9.

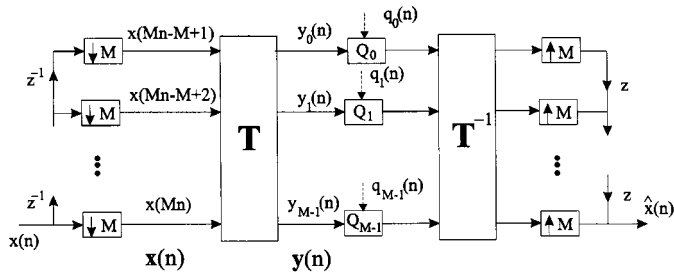


Fig. 1. Transform coder.

- 4) Its elements can be adapted as frequently as we like without affecting the PR condition. A universal transform coder without the need to send any side information can be implemented using the proposed PLT structures.
- 5) The same structures of PLT can be used for both lossy and lossless data compression.
- 6) For AR(1) input, PLT has a simple closed-form expression and can be found by inspection. Its implementation needs only  $(M - 1)$  multiplications and additions. Thus, its complexity is lower than the DCT, which has a complexity of  $\mathcal{O}(M \log M)$ . Moreover, the PLT is optimal for all AR(1) processes, unlike DCT, which is optimal only when the correlation coefficient approaches 1.

*Paper Outline:* The paper is presented as follows. Section II briefly reviews the theory of transform coding and linear prediction coding (LPC). The PLT is derived in Section III. In Section IV, two minimum noise structures are introduced, and we will show how to implement universal transform coders from the proposed structures. The application of PLT to lossless coding is discussed in Section V. In Section VI, generalization of PLT is studied. In Section VII, we consider PLT for AR(1) inputs. Some partial results and a different approach for the derivation of PLT have been reported in [18], [19].

*Notations:* Vectors and matrices will, respectively, be denoted by boldfaced lowercase and uppercase letters. An  $M \times M$  diagonal matrix with diagonal elements  $d_i$  will be expressed as  $\mathbf{D} = \text{diag}(d_0, d_1, \dots, d_{M-1})$ .

## II. PRELIMINARIES AND REVIEWS

In this section, we will first state the noise model of this paper. Then, we will briefly review various properties of the KLT and LPC. Their connection will be mentioned without proofs. Most of these results can be found in [1]–[3] and [20].

*Signal and Quantizer Models:* In this paper, we assume that the input  $x(n)$  is a zero-mean real-valued wide-sense stationary process and that its  $k$ th autocorrelation coefficients are denoted as  $r(k)$ . The quantizers  $Q$  are scalar quantizers and can be modeled as an additive noise source. We assume that for a  $b$  bit quantizer, the variance of the quantization error  $q(n)$  satisfies

$$\sigma_q^2 = c \cdot 2^{-2b} \sigma_x^2$$

where  $\sigma_x^2$  is the variance of  $x(n)$ , which is the input to the quantizer. The quantity  $c$  is a constant that depends only on the statistics of  $x(n)$ .

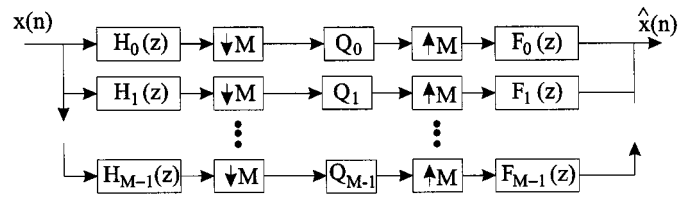


Fig. 2. Filter bank representation of the transform coder shown in Fig. 1.

### A. Transform Coders and the KLT

Consider the transform coding system in Fig. 1. Such a coding system has been studied in detail [1]–[3]. In a transform coder, the polyphase matrix  $\mathbf{T}$  is a constant matrix. Using the polyphase representation, the transform coder can be redrawn as the  $M$ -channel filter bank structure as in Fig. 2. The analysis filters  $H_k(z)$  and synthesis filters  $F_k(z)$  are, respectively, related to the polyphase matrices as

$$\begin{pmatrix} H_0(z) \\ H_1(z) \\ \vdots \\ H_{M-1}(z) \end{pmatrix} = \mathbf{T} \begin{pmatrix} z^{-M+1} \\ z^{-M+2} \\ \vdots \\ 1 \end{pmatrix} \\ (F_0(z) \ F_1(z) \ \cdots \ F_{M-1}(z)) \\ = (z^{M-1} \ z^{M-2} \ \cdots \ 1) \mathbf{T}^{-1}. \quad (1)$$

Let  $\mathbf{x}(n) = [x(Mn-M+1) \ \cdots \ x(Mn-1) \ x(Mn)]^T$  be the input vector. Then, its autocorrelation matrix is given by

$$\begin{aligned} E\{\mathbf{x}(n)\mathbf{x}^T(n)\} \\ &= \mathbf{R}_x(M) \\ &= \begin{pmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r(1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(M-1) & r(M-2) & \cdots & r(0) \end{pmatrix} \end{aligned}$$

where  $M$  indicates the dimension of the autocorrelation matrix. Many properties of autocorrelation matrix can be found in [2] and [3]. In this paper, we will assume that  $\mathbf{R}_x(M)$  is positive definite. This is, in general, true, except for the rare cases of line spectral processes. The autocorrelation matrix of the subband vector  $\mathbf{y}(n)$  is given by

$$\mathbf{R}_y(M) = \mathbf{T} \mathbf{R}_x(M) \mathbf{T}^T. \quad (2)$$

To compare the performance of different transforms, one of the commonly used measure is the coding gain. The coding gain of a transform coder is defined as ratio of the mean square error in pulse coded modulation (PCM) over that in the transform coder. For unitary transforms, the coding gain under optimal bit allocation is given by [1]

$$\mathcal{CG} = \frac{\sigma_x^2}{\prod_{k=0}^{M-1} [\sigma_{y_k}^2]^{1/M}}$$

where  $\sigma_x^2$  and  $\sigma_{y_i}^2$  are, respectively, the variances of  $x(n)$  and  $y_i(n)$ . It is well known that the unitary matrix that maximizes the coding gain is the KLT. The maximized coding gain is

$$\mathcal{CG}_{\text{KLT}}(M) = \frac{\sigma_x^2}{[\det \mathbf{R}_x(M)]^{1/M}} \quad (3)$$

where  $M$  indicates the dimension of the transform. The KLT is the unitary matrix that diagonalizes  $\mathbf{R}_x(M)$ . The columns of the KLT matrix consist of the eigenvectors of  $\mathbf{R}_x(M)$ . It can be shown that  $\mathcal{CG}_{\text{KLT}}(M) \geq 1$  with equality if and only if the autocorrelation matrix  $\mathbf{R}_x(M) = \sigma_x^2 \mathbf{I}$ . Note that the coding gain of KLT is a nondecreasing function of the dimension  $M$ . Moreover  $\mathcal{CG}_{\text{KLT}}(M) = \mathcal{CG}_{\text{KLT}}(M-1)$  if and only if  $\mathcal{CG}_{\text{KLT}}(M) = \mathcal{CG}_{\text{KLT}}(M-1) = \dots = \mathcal{CG}_{\text{KLT}}(1) = 1$  [1].

It was shown in [6] that under the assumption of high bit rate noise model, it is not a loss of generality to assume that the transform is a unitary transform. In other words, the coding gain of any transform (including nonunitary and unitary transforms) cannot be higher than that of the KLT.

### B. Linear Prediction Coding (LPC)

The LPC theory has been studied for decades, and excellent introduction to LPC can be found in [2], [3], and [20]. In an LPC problem, for a given WSS input  $x(n)$ , we want to find a filter of the form  $P_N(z) = 1 + p(1)z^{-1} + \dots + p(N)z^{-N}$  such that its output  $e_N(n)$  has a minimum variance. The filter  $P_N(z)$  is called the  $N$ th-order prediction error filter, and its output  $e_N(n)$  is the prediction error. The optimal prediction filter  $P_N(z)$  can be obtained by solving the following normal equation:

$$\mathbf{R}_x(N) \begin{pmatrix} 1 \\ p(1) \\ \vdots \\ p(N) \end{pmatrix} = \begin{pmatrix} \mathcal{E}_p(N) \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (4)$$

The above normal equation can be solved by the Levinson–Durbin fast algorithm in  $\mathcal{O}(N^2)$ . The prediction error variance  $\mathcal{E}_p(N)$  is

$$\mathcal{E}_p(N) = r(0) - \sum_{i=1}^N p(i)r(i).$$

The index  $N$  indicates the order of prediction error filter. The prediction error is a nonincreasing function of  $N$  [2]. Moreover, the prediction error variance  $\mathcal{E}_p(k)$  is related to  $\det[\mathbf{R}_x(M)]$  as

$$\det[\mathbf{R}_x(M)] = \mathcal{E}_p(0)\mathcal{E}_p(1)\cdots\mathcal{E}_p(M-1). \quad (5)$$

Furthermore, it can be shown by using the orthogonality principle [3], [20] that the corresponding prediction errors  $e_{N,\text{opt}}(n)$  satisfy the following property:

$$E\{e_{i,\text{opt}}(n)e_{j,\text{opt}}(n-k)\} = 0 \quad \text{for } 1 \leq k \leq |i-j|, \quad \text{for all } n. \quad (6)$$

In LPC, the prediction gain is a commonly used quantity to describe the effectiveness of a predictor and it is defined as

$$G_P(N) = \frac{\sigma_x^2}{\mathcal{E}_p(N)}. \quad (7)$$

In a closed-loop differential pulse code modulation (DPCM) system, the coding gain is given by the prediction gain  $G_P(N)$ .

### III. PREDICTION-BASED LOWER TRIANGULAR TRANSFORM (PLT)

In this section, we will show how to construct the PLT from a given autocorrelation matrix. Before the derivation of the PLT, we will make some definition and state a matrix decomposition lemma from the matrix theory.

*Definition [21]:* Given an  $N \times N$  matrix  $\mathbf{A}$ , its principle submatrix of dimension  $K$  (where  $K \leq N$ ) is a  $K$  by  $K$  matrix  $\mathbf{A}_K$ , with its elements  $[\mathbf{A}_K]_{ij} = [\mathbf{A}]_{ij}$  for  $0 \leq i, j \leq K-1$ . For example

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix}.$$

Then, the principle submatrices of  $\mathbf{A}$  of dimension 1, 2, 3 are, respectively

$$\mathbf{A}_1 = 1; \quad \mathbf{A}_2 = \begin{pmatrix} 1 & 2 \\ 5 & 6 \end{pmatrix}; \quad \mathbf{A}_3 = \begin{pmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{pmatrix}.$$

*Lemma 1—LU Decomposition of Matrices [21]:* Let  $\mathbf{A}$  be an  $N$  by  $N$  nonsingular matrix. Suppose that all of its principle submatrices  $\mathbf{A}_K$  are nonsingular. Then,  $\mathbf{A}$  can be written as

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$$

in which  $\mathbf{L}$  (respectively,  $\mathbf{U}$ ) is a lower (respectively, upper) triangular matrix with all diagonal entries equal to 1, and  $\mathbf{D}$  is a diagonal matrix. Moreover, the matrices  $\mathbf{L}$ ,  $\mathbf{U}$ , and  $\mathbf{D}$  are unique. In particular,  $\mathbf{D}$  is determined by

$$\det[\mathbf{D}_K] = \det[\mathbf{A}_K], \quad K = 1, \dots, N.$$

#### A. PLT

Consider the transform coder shown in Fig. 1. At the encoder, the autocorrelation matrices of the input vector  $\mathbf{x}(n)$  and the output vector  $\mathbf{y}(n)$  are related as (2). Since  $\mathbf{R}_x(M)$  is positive definite, all of its principle submatrices are positive definite as well and, therefore, nonsingular [21]. Thus, applying the LU decomposition lemma, the matrix  $\mathbf{R}_x(M)$  can be expressed as

$$\mathbf{R}_x(M) = \mathbf{L}\mathbf{D}\mathbf{U} \quad (8)$$

with the matrices  $\mathbf{L}$ ,  $\mathbf{U}$ , and  $\mathbf{D}$  defined in Lemma 1. Moreover, since  $\mathbf{R}_x(M)$  is symmetric, we can take the transpose of (8) and obtain

$$\mathbf{R}_x(M) = \mathbf{U}^T \mathbf{D} \mathbf{L}^T. \quad (9)$$

The matrices  $\mathbf{U}^T$  and  $\mathbf{L}^T$  are, respectively, lower and upper triangular. Therefore, (9) is also an LU decomposition of  $\mathbf{R}_x(M)$ . From Lemma 1, we know that the LU decomposition is unique. Thus, we conclude that

$$\mathbf{U} = \mathbf{L}^T. \quad (10)$$

Substituting (10) into (8) and simplifying the results, we can conclude that there exists a unique lower triangular matrix  $\mathbf{P} = \mathbf{L}^{-1}$  such that

$$\mathbf{D} = \text{diag}(d_0, d_1, \dots, d_{M-1}) = \mathbf{P}\mathbf{R}_x(M)\mathbf{P}^T. \quad (11)$$

The diagonal matrix  $\mathbf{D}$  is uniquely determined by

$$\prod_{i=0}^{K-1} d_i = \det[\mathbf{R}_x(K)].$$

Using the fact that  $\det[\mathbf{R}_x(K)] = \mathcal{E}_p(0)\mathcal{E}_p(1)\cdots\mathcal{E}_p(K-1)$ , we get

$$d_i = \mathcal{E}_p(i).$$

The  $i$ th entry of  $\mathbf{D}$  is the prediction error variance of an  $i$ th-order optimal predictor! Comparing the results in (2) and (11), the input autocorrelation matrix  $\mathbf{R}_x(M)$  can be diagonalized by taking the transform  $\mathbf{T}$  as the unique lower triangular matrix  $\mathbf{P}$ .

*Finding the Unique Lower Triangular Transform  $\mathbf{P}$ :* Let the lower triangular matrix  $\mathbf{P}$  be expressed as

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ p_{1,0} & 1 & 0 & \cdots & 0 \\ p_{2,0} & p_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{M-1,0} & p_{M-1,1} & p_{M-1,2} & \cdots & 1 \end{pmatrix}. \quad (12)$$

Since  $\det[\mathbf{P}] = 1$ , the inverse transform  $\mathbf{P}^{-1}$  always exists and is also lower triangular with unity diagonal elements. Let  $\mathbf{S}$  denote the inverse transform  $\mathbf{P}^{-1}$

$$\mathbf{S} = \mathbf{P}^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ s_{1,0} & 1 & 0 & \cdots & 0 \\ s_{2,0} & s_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{M-1,0} & s_{M-1,1} & s_{M-1,2} & \cdots & 1 \end{pmatrix}. \quad (13)$$

Using (1), (12) and (13), we can write the analysis and synthesis filters in Fig. 2, respectively, as

$$\begin{aligned} H_k(z) &= z^{-M+1+k} \\ &\quad \cdot (1 + p_{k,k-1}z^{-1} + p_{k,k-2}z^{-2} + \cdots + p_{k,0}z^{-k}) \\ &= z^{-M+1+k}P_k(z), \end{aligned} \quad (14)$$

$$\begin{aligned} F_k(z) &= z^{M-k-1} + s_{k+1,k}z^{M-k-2} + s_{k+2,k}z^{M-k-3} \\ &\quad + \cdots + s_{M-1,k}. \end{aligned} \quad (15)$$

From (14),  $P_k(z)$  can be viewed as a  $k$ th-order prediction error polynomial, and the output of the  $k$ th analysis filter  $H_k(z)$  is the corresponding prediction error delayed by  $(M-k-1)$  samples. If  $P_k(z)$  is optimal, then the output of  $H_k(z)$  will be  $e_{k,opt}(n-M+k+1)$ , and its variance is  $\mathcal{E}_p(k)$ . Since the decimator does not change the variance, we have  $\sigma_{y_k}^2 = \mathcal{E}_p(k)$ . Using the property in (6), we can see that  $E\{e_{k,opt}(n-M+k+1)e_{j,opt}(n-M+j+1)\} = 0$ , for  $k \neq j$ . In other words, the autocorrelation matrix  $\mathbf{R}_y(M)$  will be the diagonal matrix  $\text{diag}(\mathcal{E}_p(0), \mathcal{E}_p(1), \dots, \mathcal{E}_p(M-1))$ . From Lemma 1, we know that the lower triangular matrix with such a decorrelation property is unique. Therefore, we conclude that if  $P_k(z)$

is the  $k$ th-order optimal prediction error filter, then the corresponding matrix  $\mathbf{P}$  in (12) is the unique lower triangular matrix. We will refer to such a matrix as the PLT. The optimal  $P_k(z)$  can be obtained by solving the normal equation in (4). Using the Levinson–Durbin fast algorithm, the PLT matrix can be computed in  $\mathcal{O}(M^2)$ . Summarizing the results, we have the following theorem.

*Theorem 2:* Consider the transform coder in Fig. 1. Given any wide sense stationary input  $x(n)$ , there exists a unique lower triangular matrix  $\mathbf{P}$  of the form in (12) such that the transform coefficients  $y_k(n)$  are uncorrelated. The unique lower triangular transform can be obtained by choosing  $P_k(z) = 1 + p_{k,k-1}z^{-1} + p_{k,k-2}z^{-2} + \cdots + p_{k,0}z^{-k}$  as the  $k$ th-order optimal prediction error filter. Moreover, the autocorrelation matrix of the subband vector  $\mathbf{y}(n)$  is

$$\mathbf{R}_y(M) = \text{diag}(\mathcal{E}_p(0), \mathcal{E}_p(1), \dots, \mathcal{E}_p(M-1))$$

where  $\mathcal{E}_p(k)$  is the prediction error variance of  $P_k(z)$ .

*Complexity of the PLT:* The PLT and its inverse are both lower triangular with unity diagonal elements, and the complexity of the transform (or its inverse) is therefore only  $0.5M(M-1)$  multiplications and additions. Compared with the case of KLT, which needs  $M^2$  multiplications and  $M(M-1)$  additions, the complexity is less than one half of the KLT. In the special case of AR(1), the complexity of the PLT further reduces to  $M-1$  multiplications and additions.

*Variations of the PLT:* In the previous discussion, we have derived the PLT for the vector input  $\mathbf{x}(n) = [x(Mn-M+1) \cdots x(Mn)]^T$ . That means that the polyphase components  $x(Mn-i)$  are arranged in an ascending order. We can also permute these polyphase components so that the new input vector  $\mathbf{x}'(n) = \mathcal{P}\mathbf{x}(n)$ , where  $\mathcal{P}$  is a permutation matrix. In this case, we can design PLT for the new autocorrelation matrix

$$\mathbf{R}_{x'}(M) = \mathcal{P}\mathbf{R}_x(M)\mathcal{P}. \quad (16)$$

In this case, the optimal transform (PLT) can be obtained by using the orthogonality principle. As we will show in the next section, the coding gains of the new PLT for all permutation matrices  $\mathcal{P}$  are identical. Although their coding performance is the same, some permutation matrices can result in PLT with lower implementational cost. If the permutation matrix is chosen judiciously, some of the coefficients can be made symmetric. To explain this, take  $M=3$ . If the input vector is taken as  $\mathbf{x}'(n) = [x(3n-2)x(3n)x(3n-1)]^T$ , then the corresponding PLT will have the form

$$\mathbf{P}' = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \beta & \beta & 1 \end{pmatrix}$$

for some  $\alpha$  and  $\beta$ . To implement  $\mathbf{P}'$ , we need only two multiplications instead of three. Using the orthogonality principle, one can verify that the transform coefficients are uncorrelated. That is,  $\mathbf{R}_{y'}(M) = E\{\mathbf{y}'(n)\mathbf{y}'^T(n)\} = \text{diag}(\sigma_{y_0}^2, \dots, \sigma_{y_{M-1}}^2)$ .

### B. Implementation of PLT Using Ladder Structures

The PLT has a structurally PR implementation using the ladder structure. In such an implementation, the filter bank

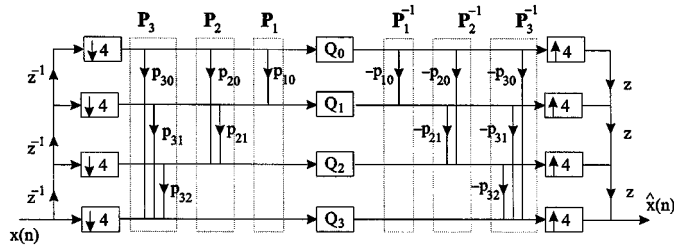


Fig. 3. Ladder-based implementation of the PLT coder.

continues to have PR even when all the multipliers in the structure are quantized to a finite precision. In the following, we will derive the ladder structure for PLT.

Note that any lower triangular matrix  $\mathbf{P}$  of the form (12) can be decomposed as

$$\mathbf{P} = \mathbf{P}_1 \mathbf{P}_2 \cdots \mathbf{P}_{M-1}$$

where

$$\mathbf{P}_k = \begin{pmatrix} \mathbf{I}_k & \mathbf{0}_{M-k-1 \times k} \\ p_{k,0} \cdots p_{k,k-1} & 1 \ 0 \ \cdots \ 0 \\ \mathbf{0}_{k \times M-k-1} & \mathbf{I}_{M-k-1} \end{pmatrix}. \quad (17)$$

The elementary matrix  $\mathbf{P}_k$  has only one nontrivial row formed by coefficients of the  $k$ th-order prediction error filter. The inverses of these elementary matrices are very simple and can be found by inspection. To be more precise, the inverse can be obtained by replacing the nontrivial elements  $p_{k,i}$  in (17) with  $-p_{k,i}$ . That is

$$\mathbf{P}_k^{-1} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0}_{M-k-1 \times k} \\ -p_{k,0} \cdots -p_{k,k-1} & 1 \ 0 \ \cdots \ 0 \\ \mathbf{0}_{k \times M-k-1} & \mathbf{I}_{M-k-1} \end{pmatrix}. \quad (18)$$

From (17) and (18), we see that both the transform  $\mathbf{P}$  and inverse transform  $\mathbf{P}^{-1}$  can be factorized into  $(M-1)$  ladder sections. The implementation based on these factorized forms are shown in Fig. 3 for  $M=4$ . The ladder-based implementation has the same complexity as direct implementation of  $\mathbf{P}$ . Both the encoder and decoder have the same set of multipliers. Therefore, even when these multipliers are implemented in finite precision, PR is still preserved.

#### IV. MINIMUM NOISE STRUCTURES FOR PLT

Recall that the PLT is a nonunitary matrix, and so is its inverse. Hence, the PLT coder does not have the energy preservation property. In general, the quantization noise generated in the subbands will be amplified at the decoder. To study how the noise is amplified, we assume that the quantization noise  $q_k(n)$  in different subband is uncorrelated. That is, we assume that  $E\{q_k(n)q_{k+i}(n)\} = \mathbf{R}_q = \text{diag}(\sigma_{q_0}^2, \sigma_{q_1}^2, \dots, \sigma_{q_{M-1}}^2)$ . Under this assumption, we can show that the average output noise variance is given by

$$\sigma_{q_{\text{out}}}^2 = \frac{1}{M} \sum_{k=0}^{M-1} \|F_k\|_2^2 \sigma_{q_k}^2$$

where  $\|F_k\|_2^2$  is the two norm of the synthesis filter  $F_k(z)$  in (15), and it is given by  $\|F_k\|_2^2 = 1 + \sum_{i=k+1}^{M-1} s_{i,k}^2$ . Therefore, the noise gain is always greater than one, unless the matrix  $\mathbf{S}$  is the identity matrix. We will call the structure in Fig. 3 the

nonminimum noise structure for the PLT. To understand why the noise gain is larger than one, let us consider Fig. 3. The inputs to the multipliers  $p_{i,j}$  at the encoder are the unquantized data, whereas the inputs to the multipliers  $p_{i,j}$  at the decoder are the quantized data. This means that the predictors at the encoder use unquantized data as their observations, whereas the predictors at the decoder use the quantized data. It is this mismatch that causes the noise amplification.

*Coding Gain for Non Minimum Noise Structure:* Using the noise model defined in Section II and applying the optimal bit allocation, one can show that the coding gain of the PLT is

$$\begin{aligned} CG_{\text{PLT, NON}}(M) &= \frac{1}{\prod_{i=0}^{M-1} [\|F_k\|_2^2]^{1/M}} \times \frac{\sigma_x^2}{[\det \mathbf{R}(M)]^{1/M}} \\ &= \frac{1}{\prod_{i=0}^{M-1} [\|F_k\|_2^2]^{1/M}} \times CG_{\text{KLT}}(M). \end{aligned}$$

Since all  $\|F_k\|_2^2 \geq 1$ , we conclude that  $CG_{\text{PLT, NON}}(M) \leq CG_{\text{KLT}}(M)$  with equality if and only if  $\mathbf{R}_x(M) = \sigma_x^2 \mathbf{I}$ . Due to the noise amplification, the quantity  $CG_{\text{PLT, NON}}(M)$  is not guaranteed to be greater than unity. In the rest of this section, we will derive two minimum noise structures that have the unity noise gain property.

#### A. MINLAB(I) Structure for PLT

Note that the inverse transform  $\mathbf{S}$  in (13) is also lower triangular. Thus, we can factorize  $\mathbf{S}$  as

$$\mathbf{S} = \mathbf{S}_1 \mathbf{S}_2 \cdots \mathbf{S}_{M-1}$$

where

$$\mathbf{S}_k = \begin{pmatrix} \mathbf{I}_k & \mathbf{0}_{M-k-1 \times k} \\ s_{k,0} \cdots s_{k,k-1} & 1 \ 0 \ \cdots \ 0 \\ \mathbf{0}_{k \times M-k-1} & \mathbf{I}_{M-k-1} \end{pmatrix}. \quad (19)$$

Using the above equation, the transform  $\mathbf{P}$  can be factorized as

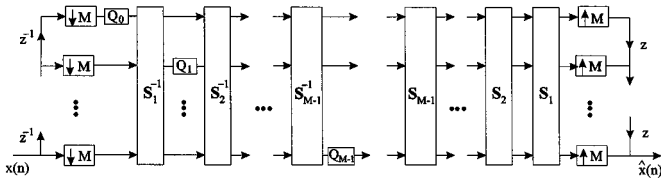
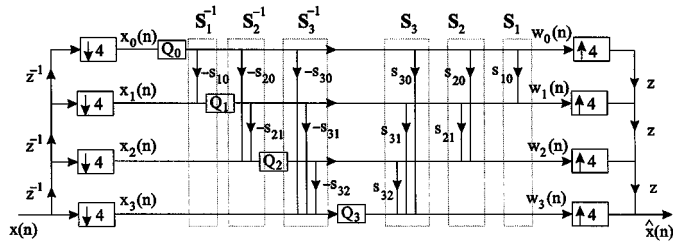
$$\mathbf{P} = \mathbf{S}_{M-1}^{-1} \mathbf{S}_{M-2}^{-1} \cdots \mathbf{S}_1^{-1} \quad (20)$$

where  $\mathbf{S}_k^{-1}$  can be obtained by simply replacing the multipliers  $s_{k,i}$  in (19) by  $-s_{k,i}$ . From (19) and (20), we can obtain a ladder-based implementation that is different from Fig. 3. Such a ladder structure will also be structurally PR, but both the encoder and decoder have the same set of multipliers  $s_{k,i}$ . Provided that  $s_{i,k}$  are quantized to the same value at the encoder and decoder, we continue to have the PR property.

Using (20), the minimum noise structure for the transform can be implemented as Fig. 4. Since  $\mathbf{S}_k^{-1}$  has the identity matrix  $\mathbf{I}_k$  on its top-left corner, the output vector of the encoder in Fig. 4 contains only quantized values. To see why the minimum noise structure has the unity noise gain property, we take  $M=4$ . The minimum noise structure for  $4 \times 4$  PLT is shown in Fig. 5. From the figure, it is not difficult to verify that

$$w_i(n) - x_i(n) = q_i(n)$$

for  $i=0, 1, 2, 3$ . The noise gain for the PLT is unity even though the transform is nonunitary. Note that we do not make

Fig. 4. MINLAB(I) structure for  $M \times M$  PLT coder.Fig. 5. Minimum noise structure MINLAB(I) for  $4 \times 4$  PLT coder.

any assumptions on the quantization noise  $q_i(n)$ . This unity noise gain property holds even when  $q_i(n)$  noises are correlated and colored. In general, one can show that the minimum noise structure in Fig. 4 for  $M \times M$  PLT has the same unity noise gain property. In the rest of this paper, we will refer to Fig. 4 as the MINLAB(I) structure for PLT.

### B. MINLAB(II) for PLT

We can also modify the ladder implementation based on factorization of  $\mathbf{P}$  in (17) to obtain a different minimum noise structure. To avoid the mismatch of observations in the ladder-based structure in Fig. 3, one can modify the structure so that the inputs to the multipliers  $p_{k,i}$  at the encoder are quantized data instead of the original unquantized values. The encoder of the modified structure for  $M = 4$  case is shown in Fig. 6, and the decoder is the same as Fig. 3. From the figure, one can verify that the structure has the unity noise gain property. For the same reasoning as in MINLAB(I) case, this property holds even for correlated and colored quantization noise. The implementation in Fig. 6 will be referred to as MINLAB(II) structure for PLT.

### C. Complexity of the MINLAB Structures

For MINLAB(I) in Fig. 4, the structure has the same number of multiplications and additions as the nonminimum noise structure in Fig. 3. However, since all the multipliers are  $s_{k,i}$  (not  $p_{k,i}$ ), we need to invert the lower triangular matrix  $\mathbf{P}$  to obtain these parameters. If Gaussian elimination method is used, we need  $M(M+1)(M+2)/6$  multiplications and additions to invert an  $M \times M$  lower triangular matrix. On the other hand, the MINLAB(II) structure requires an extra  $(M-1)$  adders compare to the nonminimum noise structure. The multipliers in MINLAB(II) are  $p_{k,i}$ , and thus, no matrix inversion is needed. Although the two MINLAB structures have the same coding gain, their complexities are not the same. For an input of length  $L$ , the MINLAB(I) structure has an overhead of  $M(M+1)(M+2)/6$  multiplications and additions, whereas the MINLAB(II) structure has an extra  $(M-1)L/M$  additions. Therefore, when  $L \gg M$ , MINLAB(I) is preferred; otherwise, MINLAB(II) is preferred.

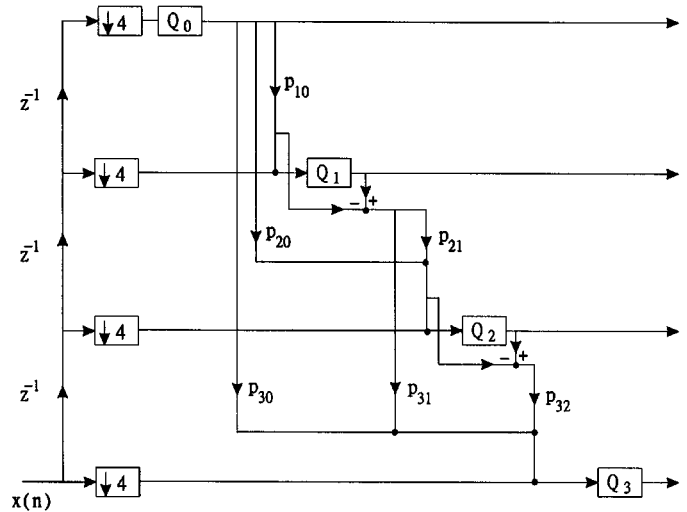


Fig. 6. Encoder for PLT MINLAB(II) structure. The decoder is the same as Fig. 3.

### D. Coding Gain of PLT Using MINLAB Implementations

The two MINLAB structures have the unity noise gain property. Therefore, for a fixed average bit rate  $b = 1/M \sum_{k=0}^{M-1} b_k$ , the average output variance can be expressed as

$$\begin{aligned} \sigma_{q_{out}}^2 &= \frac{1}{M} \sum_{k=0}^{M-1} \sigma_{q_i}^2 \\ &= \frac{1}{M} \sum_{k=0}^{M-1} c 2^{-2b_k} \mathcal{E}_p(k) \end{aligned}$$

where we have used the fact that the  $k$ th subband signal variance is  $\mathcal{E}_p(k)$ . Applying the arithmetic mean geometric mean inequality to the above equation, we get

$$\sigma_{q_{out}}^2 \geq c 2^{-2b} \prod_{i=0}^{M-1} [\mathcal{E}_p(i)]^{1/M}$$

with equality if and only if the bits are allocated as

$$b_k = b + 0.5 \log_2 \mathcal{E}_p(k) - 0.5 \log_2 \prod_{i=0}^{M-1} [\mathcal{E}_p(i)]^{1/M}.$$

From the above derivation, we see that the average output noise variance is minimized if all quantizers have the same noise variance. Therefore, the equal stepsize rule is also optimal, and entropy coding can be used to encode the outputs of  $Q_k$ .

Compared with the error variance in a PCM system, the coding gain of PLT is given by

$$\mathcal{CG}_{PLT,MIN}(M) = \frac{\sigma_x^2}{\prod_{i=0}^{M-1} [\mathcal{E}_p(i)]^{1/M}}$$

where the subscript  $PLT, MIN$  indicates that the coding gain is for the minimum noise structures of PLT. Using (3) and (5), we conclude that the coding gain of PLT is

$$\mathcal{CG}_{PLT,MIN}(M) = \frac{\sigma_x^2}{[\det \mathbf{R}_x(M)]^{1/M}} = \mathcal{CG}_{KLT}(M). \quad (21)$$

Therefore, the PLT coders with minimum noise structures have the same coding gain as KLT. Using (5) and (7), one can express the coding gain of PLT in terms of the prediction gain

$$\mathcal{CG}_{PLT, MIN}(M) = [G_p(M-1) \cdots G_p(1)G_p(0)]^{1/M}. \quad (22)$$

*Remark:* As we have mentioned at the end of Section III-A, we can also derive PLT for the input vector  $\mathbf{x}'(n) = \mathcal{P}[x(Mn-M+1) \cdots x(Mn)]^T$ . In this case, the autocorrelation matrix  $\mathbf{R}_{\mathbf{x}'}(M)$  for  $\mathbf{x}'(n)$  is related to  $\mathbf{R}_{\mathbf{x}}(M)$  as (16). Since  $\prod_{i=0}^{M-1} \sigma_{y_i}^2 = \det \mathbf{R}_{\mathbf{y}'}(M) = \det \mathbf{R}_{\mathbf{x}'}(M) = \det \mathbf{R}_{\mathbf{x}}(M)$ , we conclude from (21) that for all permutation matrices  $\mathcal{P}$ , the coding gain is the same.

*Comparison with DPCM:* The DPCM is also a prediction-based coding system. For a closed loop DPCM with a prediction filter of order  $(M-1)$ , the coding gain is given by the prediction gain  $G_p(M-1)$ . Since the prediction gain is an increasing function of the filter order (except for AR processes where the gain saturates), we can conclude from (22) that  $\mathcal{CG}_{DPCM}(M-1) > \mathcal{CG}_{PLT, MIN}(M)$ . Even though the coding gain of a PLT is less than that of a DPCM, it has other advantages.

- 1) Unlike DPCM, the PLT involves only FIR filtering in the reconstruction process. Therefore, any error occurring in the transmission or storage will not be propagated.
- 2) In PLT, the relationship of  $y_i(n)$  in different channels can be exploited for further compression, e.g., zerotree algorithm [22].
- 3) *The computational complexity:* The PLT requires  $0.5(M-1)$  multiplications and additions per input sample, whereas the DPCM encoder (or decoder) of the same order needs  $(M-1)$  multiplications and additions per input sample.

*Effect of Quantization on the Prediction Gain:* At very low bit rate coding, the quantized data  $Q[x(n)] = x(n) + q(n)$  can be very different from  $x(n)$ . If the SNR decreases, the accuracy of the estimate by using these quantized data will decrease. The prediction gain will decrease. Therefore, like other prediction-based coding methods, the coding gain of PLT will decrease when the SNR decreases.

*Universal Transform Coder:* Since the MINLAB structures are structurally PR, we can adapt the multipliers as frequently without affecting the PR property. The statistics of the input can be adaptively estimated from the quantized data, and this information can be used to update the prediction error polynomials. Since the estimation is based on the quantized data, there is no need to send any side information to the decoder. Given any input signal, we can initialize the PLT as  $\mathbf{P}^{(0)} = \mathbf{I}$ . After each input vector  $\mathbf{x}(n)$  is encoded with  $\mathbf{P}^{(n)}$ , the statistics can be updated, and the transform  $\mathbf{P}^{(n+1)}$  can be computed in  $\mathcal{O}(M^2)$  using the Levinson–Durbin fast algorithm. After a few iterations, if the statistics of the input do not vary too fast, the rate of adaptation can be reduced. In addition, the transform can be updated only after a number of input vectors are encoded. For the implementation of universal coder, MINLAB(II) structure is preferred because MINLAB(I) structure needs to invert  $\mathbf{P}^{(n)}$  for each  $n$ .

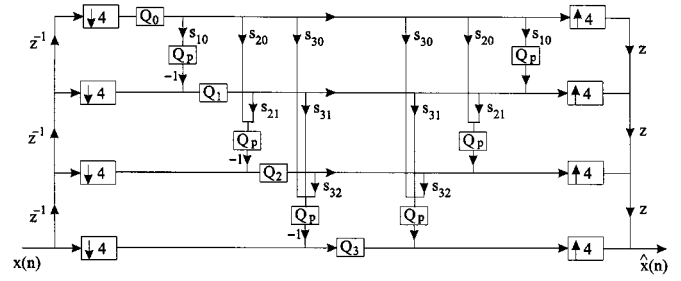


Fig. 7. Lossy/lossless PLT coder.

One can also use adaptive algorithms such as the least mean square (LMS) method to update the  $(M-1)$  different predictors. In this case, there is no need to estimate the statistics, and the complexity will be  $\mathcal{O}(M^2)$ .

## V. PLT FOR LOSSLESS DATA COMPRESSION

In many applications, it is desired that a lossy coding system becomes lossless when a sufficient bit rate is available. Since the multipliers of KLT are real numbers, in practice, they have to be quantized. In general, the KLT with quantized multipliers will not have the PR property. Therefore, the KLT, in general, cannot be used for lossless coding.

On the other hand, the two MINLAB structures introduced in previous section can be implemented for both lossy and lossless coding after some minor modifications. To see this, assume that the input values  $x(n)$  are integers. Take the MINLAB(I) structure in Fig. 5 as an example. If a quantizer  $Q_p$  is cascaded after all of the predictors as shown in Fig. 7 and its stepsize is set as  $\Delta_p = 1$ , then the PR property continues to hold. The quantizer  $Q_p$  in Fig. 7 can be roundoff, truncation, or ceiling quantizer. Recall that the PLT coder is optimal when the equal stepsize rule is applied. Therefore, we can set the stepsize  $\Delta_k$  of quantizer  $Q_k$  to the same value, and entropy coding can be used to encode the quantized subband signals. If all  $\Delta_k = n > 1$ , then we have a lossy PLT coder. If all  $\Delta_k = 1$ , then the PLT coder becomes lossless. Therefore, we can implement both lossy and lossless coding with the same PLT coder by simply adjusting the stepsizes  $\Delta_k$ . Similarly, one can modify the MINLAB(II) structure in Fig. 6 to obtain a lossy/lossless coder.

## VI. GENERALIZED PLT

The transform coder has a constant polyphase matrix  $\mathbf{T}$ . It is a special class of subband coding, where the polyphase matrix is a polynomial matrix. Since the polyphase matrix is constant, the PLT discussed in an earlier section can only exploit the correlation of data within each input vector. Therefore, its performance is limited by its transform size  $M$ . In order to exploit the correlation among input vectors, one can replace the entries  $p_{k,i}$  with the more general FIR filters  $p_{k,i}(z)$

$$P_k(z) = 1 + p_{k,k-1}(z^M)z^{-1} + \cdots + p_{k,0}(z^M)z^{-k}. \quad (23)$$

For this generalized PLT, the ladder-based structure in Fig. 3 continues to be structurally PR. Moreover, the two MINLAB

structures continue to enjoy the unity noise gain property. Therefore, the average output noise variance after optimal bit allocation is

$$\sigma_{q_{\text{out}}}^2 \geq c2^{-2b} \prod_{i=0}^{M-1} [\sigma_{y_i}^2]^{1/M}$$

where  $\sigma_{y_i}^2$  is the variance of the  $i$ th subband signal  $y_i(n)$ . To maximize the coding gain,  $p_{k,i}(z)$  should be chosen such that the energy of the subband signals is minimized. This is the well-known linear estimation problem, and the solution can be obtained using the orthogonality principle. As the estimation error decreases when more samples are used in the estimation, the coding gain increases when the length of  $P_k(z)$  increases. If all  $p_{k,i}(z)$  are causal, the system delay is still  $(M-1)$ . Therefore, we would have a coder with better performance without increasing the system delay.

#### A. Generalized PLT with Interpolation

If the  $p_{k,i}(z)$  in (23) is taken as a noncausal polynomial, that is,  $p_{k,i}(z) = \sum_{n=-N}^{N-1} \alpha(n)z^{-n}$ , then interpolation (instead of extrapolation) value is used as the estimate. In a generalized PLT, the input is partitioned into  $M$  nonoverlapped polyphase components  $x(Mn-i)$  for  $0 \leq i \leq M-1$ . We are estimating  $x(Mn-k)$  from  $x(Mn-i)$  for  $i < k$ . Therefore, noncausal estimation can be implemented. In this case, the estimation error is smaller, and the resulting coding gain is higher than the PLT.

One special case of such a noncausal PLT is the hierarchical interpolation (HINT) coder. The HINT coder has the advantage that its complexity is very low. The HINT coder has been applied to lossless compression [11], [12]. However, since the encoder uses unquantized data for its estimation while the decoder uses quantized data for its estimation, the structure used in [11] and [12] does not have the unity noise gain property. Therefore, in general, the HINT coder does not give a satisfactory result when applied to lossy compression. To explain how we can get a HINT coder with MINLAB structures, let the number of channel  $M = 2^3$ . In an eight-channel HINT coder, the input is partitioned into four groups:

$$\begin{aligned} \diamond &= \{x(n): n = 8k, k \in \text{integer}\}; \\ \star &= \{x(n): n = 8k + 4, k \in \text{integer}\}; \\ \circ &= \{x(n): n = 4k + 2, k \in \text{integer}\}; \\ \bullet &= \{x(n): n = 2k + 1, k \in \text{integer}\}. \end{aligned}$$

These groups are shown in Fig. 8. In a MINLAB HINT coder, the samples in Group  $\diamond$  are quantized directly. The samples in Group  $\star$  are first estimated from the two nearest *quantized* samples in Group  $\diamond$ , and then the estimation error is quantized. Similarly samples in other groups are estimated from the two nearest samples in *previously quantized* groups and the estimation error is coded. The complexity of MINLAB HINT is very low. Each estimation takes only one multiplication because of symmetry. To implement the above eight-channel system, the encoder needs only seven multiplications for encoding eight input samples. In general, for an  $M$ -channel MINLAB HINT coder, the encoder (or decoder) needs only  $(M-1)/M$  multiplications per input sample.

Fig. 8. Eight-channel hint coder partitions the input into four groups  $\diamond$ ,  $\star$ ,  $\circ$ , and  $\bullet$ .

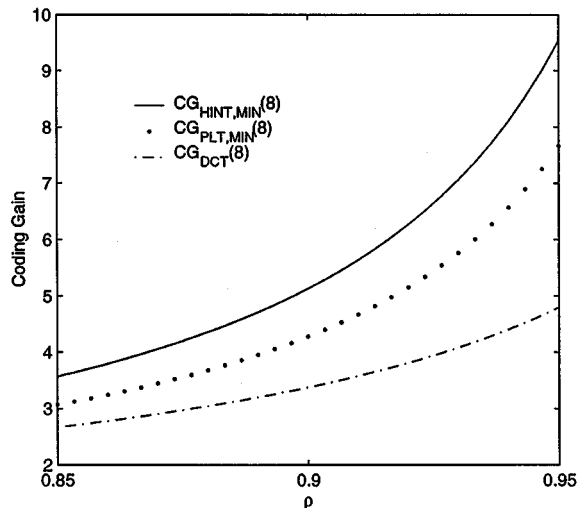


Fig. 9. Comparison of eight-channel MINLAB HINT, PLT, and DCT coders for AR(1) with correlation  $\rho$ .

#### VII. PLT FOR AR(1) INPUTS

If the input is an AR(1) process with correlation  $\rho$ , then all the prediction error polynomials  $P_k(z)$  in (14) will have the same form  $(1 - \rho z^{-1})$ . The PLT in this case has the following closed form:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}. \quad (24)$$

Once  $\rho$  is known, we can find  $\mathbf{P}$  by inspection; no computation is needed. Therefore, the universal optimal PLT coder introduced in Section V becomes very simple, and we need to estimate only one parameter  $\rho$ . The coding gain in this case becomes

$$CG_{\text{PLT,MIN}}(M) = \left[ \frac{1}{1 - \rho^2} \right]^{(M-1)/M}.$$

As  $M$  is large, the above gain approaches the prediction gain of a DPCM coder.

In addition, note that the transform in (24) is almost independent of the input signal. An  $M \times M$  PLT for AR(1) process needs only  $(M-1)$  multiplications and additions. Thus, its complexity is lower than the DCT, which has a complexity of  $\mathcal{O}(M \log M)$ . Moreover, the PLT in (24) is optimal for all AR(1) processes, unlike the DCT, which is optimal only when  $\rho$  approaches 1.

*Comparison of DCT, PLT, and Generalized PLT Coders:* To compare the performances of these three coders, we use AR(1) signal as an input. As MINLAB HINT coder has a low complexity, we choose this special case to demonstrate the performance of generalized PLT with interpolation. One can show that



the coding gain of an  $2^L$ -channel MINLAB HINT coder is given by

$$\mathcal{CG}_{\text{HINT, MIN}}(2^L) = \prod_{k=1}^L \left( \frac{1 + \rho^{2^k}}{1 - \rho^{2^k}} \right)^{2^{-k}}.$$

For  $M = 2^L = 8$  and  $0.85 < \rho < 0.95$ , the coding gains of these coders are plotted in Fig. 9. One can see that as  $\rho$  approaches 1, the HINT coder is much better than the PLT, whereas the PLT is much better than the DCT.

### VIII. CONCLUDING REMARKS

In this paper, we have introduced a new nonunitary transform that has the same coding performance as the KLT. The proposed PLT coder has a lower design and implementational cost. In addition, the PLT can be applied to implement universal coders and lossy/lossless coders. Moreover, the PLT can be generalized to the overlapped transform case. The generalized PLT includes HINT coder as a special case. For AR(1) process, both PLT and HINT coders have much higher coding gain than the DCT. All these features and merits make the PLT an invaluable tool for signal compression.

### REFERENCES

- [1] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [2] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [3] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.
- [4] V. Goyal, J. Zhuang, M. Vetterli, and C. Chan, "Transform coding using adaptive bases and quantization," in *Proc. IEEE Int. Conf. Image Process.*, 1996, pp. 365–368.
- [5] M. Effros and P. A. Chou, "Weighted universal transform coding: Universal image compression with the Karhunen–Loeve transform," in *Proc. IEEE Int. Conf. Image Process.*, 1995, pp. 61–64.
- [6] P. P. Vaidyanathan, "Theory of optimal orthonormal subband coders," *IEEE Trans. Signal Process.*, vol. 46, pp. 1528–1543, June 1998.
- [7] M. K. Tsatsanis and G. B. Giannakis, "Principle component filter banks for optimal multiresolution analysis," *IEEE Trans. Signal Processing*, pp. 1766–1777, Aug. 1995.
- [8] A. Kirac and P. P. Vaidyanathan, "Theory and design of optimal FIR compaction filters," *IEEE Trans. Signal Processing*, vol. 46, pp. 903–919, Apr. 1998.
- [9] J. Tuqan and P. P. Vaidyanathan, "Globally optimal two channel FIR orthonormal filter banks adapted to the input signal statistics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1998.
- [10] S. Akkarakaran and P. P. Vaidyanathan, "On optimization of filter banks with denoising applications," in *Proc. IEEE ISCAS*, Orlando, FL, June 1999.
- [11] P. Roos, A. Viergever, and M. C. A. van Dijke, "Reversible intraframe compression of medical images," *IEEE Trans. Med. Imag.*, vol. 7, pp. 328–336, Sept. 1988.

- [12] G. R. Kuduvali and R. M. Rangayyan, "Performance analysis of reversible image compression techniques for high-resolution digital telerradiology," *IEEE Trans. Med. Imag.*, vol. 11, pp. 430–445, Sept. 1992.
- [13] A. Said and W. A. Pearlman, "An image multiresolution representation for lossless and lossy compression," *IEEE Trans. Image Processing*, vol. 5, pp. 1303–1311, Sept. 1996.
- [14] W. J. Ho and W. T. Chang, "Multiresolution interpolative DPCM for data compression," *Proc. SPIE Visual Commun. Image Process.*, 1995.
- [15] A. R. Calderbank, I. Daubechies, W. Sweldens, and B. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Proc. IEEE Int. Conf. Image Process.*, 1997, p. 596.
- [16] M. Iwahashi, S. Fukuma, and N. Kambayashi, "Lossless coding of still images with four channel prediction," in *Proc. IEEE Int. Conf. Image Process.*, 1997, p. 266.
- [17] S. M. Phoong and Y. P. Lin, "MINLAB: Minimum noise structure for ladder-based biorthogonal coders", submitted for publication.
- [18] —, "PLT versus KLT," in *Proc. IEEE Int. Symp. Circ. Syst.*, May 1999.
- [19] —, "PLT: A New class of low complexity optimal transform coders," in *Proc. Int. Symp. Opt. Sci., Eng., Instrum. (SPIE)*, July 1999.
- [20] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 61, pp. 561–580, 1975.
- [21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.
- [22] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets," *IEEE Trans. Signal Processing*, pp. 3445–3462, Dec. 1993.



**See-May Phoong** (M'96) was born in Johor, Malaysia, in 1968. He received the B.S. degree in electrical engineering from the National Taiwan University (NTU), Taipei, Taiwan, R.O.C., in 1991 and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology (Caltech), Pasadena, in 1992 and 1996, respectively.

He was with the faculty of the Department of Electronic and Electrical Engineering, Nanyang Technological University, Singapore, from September 1996 to September 1997. Since September 1997, he has

been an Assistant Professor with the Institute of Communication Engineering and Electrical Engineering, NTU. His interests include signal compression, transform coding, and filter banks and their applications to communication.

Dr. Phoong is the recipient of the 1997 Wilts Prize from Caltech for outstanding independent research in electrical engineering.



**Yuan-Pei Lin** (S'93–M'97) was born in Taipei, Taiwan, R.O.C., in 1970. She received the B.S. degree in control engineering from the National Chiao-Tung University (NCTU), Hsinchu, Taiwan, in 1992 and the M.S. and the Ph.D. degrees, both in electrical engineering, from the California Institute of Technology, Pasadena, in 1993 and 1997, respectively.

She joined the Department of Electrical and Control Engineering of NCTU in 1997. Her research interests include multirate filter banks, wavelets, and applications to communication systems. She is currently an Associate Editor of *Multidimensional Systems and Signal Processing*.