Fast Multiplierless Approximation of the DCT with the Lifting Scheme

Jie Liang and Trac D. Tran

Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218
E-mail: jieliang@jhu.edu, trac@jhu.edu

ABSTRACT

In this paper, we present a systematic approach to design two families of fast multiplierless approximations of the DCT with the lifting scheme, based on two kinds of factorizations of the DCT matrix with Givens rotations. A scaled lifting structure is proposed to reduce the complexity of the transform. Analytical values of all lifting parameters are derived, from which dyadic values with different accuracies can be obtained through finite-length approximations. This enables low-cost and fast implementations with only shift and addition operations. Besides, a sensitivity analysis is developed for the scaled lifting structure, which shows that for certain rotation angles, a permuted version of it is more robust to truncation errors. Numerous approximation examples with different complexities are presented for the 8-point and 16-point DCT. As the complexity increases, more accurate approximation of the floating DCT can be obtained in terms of coding gains, frequency responses, and mean square errors of DCT coefficients. Hence the lifting-based fast transform can be easily tailored to meet the demands of different applications, making it suitable for hardware and software implementations in real-time and mobile computing applications.

Keywords: DCT, lifting scheme, multiplierless.

1. INTRODUCTION

The Discrete Cosine Transform (DCT)[1,2] is a robust approximation of the optimal Karhunen-Loève transform (KLT) for the data with $AR(1)$ model and large correlation coefficient. It has satisfactory performance in term of energy compaction capability. Besides, many fast DCT algorithms with efficient hardware and software implementations have been proposed in the last twenty years. Therefore the DCT is extremely useful in image and video processing, and it has become the heart of many international standards such as JPEG, H.263 and MPEG.[3–5]

Most of the available fast DCT algorithms can be classified into two categories[6,7]: (i) fast algorithms taking advantage of the relationships between the DCT and various existing fast transforms such as the FFT,[1,8,9] Walsh-Hadamard transform (WHT),[10] and discrete Hartley transform (DHT)[11]; (ii) fast algorithms based on the sparse factorizations of the DCT matrix,[12,13] including some recursive methods..[14,15]

The theoretical lower bound on the number of multiplications required in the 1D 8-point DCT has been proven to be 11.[16,17] In this sense, the method proposed by Loeffler *et. al.*,[13] with 11 multiplications and 29 additions, is the most efficient solution. However, in image and video processing, quantization is often required to compress the data. In these circumstances, more efficient DCT implementations become possible if we can incorporate some multiplications of the DCT into the quantization steps.[5] For example, after proper scaling operations, 8 of the 13 multiplications in Arai's method can be pushed to the end of the 8-point DCT transform and combined with the quantization, leading to a fast DCT implementation with only 5 multiplications.[9,3]

However, these fast algorithms still need floating-point or fixed-point multiplications, which is costly in both hardware and software implementations, especially for hand-held devices. In this paper, we will present two families of multiplierless approximations of the DCT with the lifting scheme.

The lifting scheme[18,19] is a new tool for constructing wavelets and wavelet transforms. It enables flexible biorthogonal transform, and isolates the degrees of freedom remaining after fixing the biorthogonal structure. One then has full control over these degrees of freedom to customize the wavelet design for different applications. It also leads to a faster implementation of the wavelet transform, and can map integers to integers with perfect invertability.

It has been proven that any orthogonal filter bank can be decomposed into delay elements and Givens rotations by the lattice factorizations.[20] Daubechies and Sweldens[19] showed further that each Givens rotation can be represented by three lifting steps. It thus follows that any orthonormal filter bank can be written in terms of lifting steps. This also holds for the DCT, since it is an orthogonal transform and the factorization of DCT with Givens rotation has been well developed.[12,13]

The first application of the lifting scheme in the DCT, dubbed binDCT, was proposed by Tran,[21–23] where it is noted that among all the fast DCT algorithms, the sparse factorizations of the DCT with Givens rotations are most suitable for the applications of the lifting scheme. In particular, Chen's factorization[12] is used to obtain lifting-based DCT. The lifting parameters are first selected by an optimization program, aiming mainly at high coding gain that closely approximates the floating DCT. These parameters are then approximated by hardware-friendly dyadic values to enable fast implementation with only binary shifts and additions. To further reduce the complexity, a scaled lifting structure is proposed, in which a Givens rotation is replaced by only two lifting steps, instead of the standard three liftings.[19] The scaling factors resulted by this new structure can be absorbed in the quantization.

However, in our previous work,[21–23] the binDCT parameters were obtained mainly by an optimization program, which is not flexible in adjusting parameters when different accuracies of approximations of the DCT are desired. In this paper, we will show that the analytical solutions for all lifting parameters, including the scaled liftings and the corresponding scaling factors can be derived explicitly. Hence different truncations can be applied to these analytical values to obtain binDCTs with different accuracies. In particular, the analytical values of the scaling factors play a crucial role in maintaining the compatability between the binDCT coefficients and the standard DCT coefficients.

In our previous results, a permuted version of the scaled lifting structure was used for certain rotation angles. Here we will develop a sensitivity analysis, which shows that for certain rotation angles, the flipped lifting architecture is more robust to the truncation errors than the normal one. This validates our previous results and provides a useful guideline for the design of the binDCT.

In addition to Chen's DCT factorization, the binDCT design method is also applied to the Loeffler's factorization, which leads to another family of efficient binDCT solutions. Moreover, a 16-point binDCT family is also proposed, based on the Loeffler's factorization for 16-point 1D DCT.

One advantage provided by the new design approach is that we can easily adjust the lifting parameters to obtain different trade-off between the complexity and the performance, measured by the coding gain of the resulted binDCT and the mean square error (MSE) between the binDCT coefficients and the floating DCT coefficients. Hence the proposed binDCT families are suitable for different software and hardware implementations.

2. PERFORMANCE MEASURES

In this section, we define some criteria used in measuring and evaluating the performance of the proposed fast transforms.

2.1. Coding Gain

Coding gain is one of the most important factors to be considered for a transform to be used in image and video compression applications. A transform with higher coding gain compacts more energy into a fewer number of coefficients. As a result, higher objective performances such as PSNR would be achieved after quantizations. Since the coding gain of the DCT approximates the optimal KLT closely, it is desired that the binDCT have similar coding gain to that of the floating DCT.

The coding gain is defined as:[24–26]

$$C_g \;=\; 10\log_{10}\,\frac{\sigma_x^2}{\left(\prod_{i=0}^{M-1}\sigma_{x_i}^2\,||f_i||^2\right)^{\frac{1}{M}}}, \tag{1}$$

where M is the number of subbands, σ_x^2 the variance of the input, $\sigma_{x_i}^2$ the variance of the i-th subband, and $||f_i||^2$ is the norm of the i-th synthesis filter. Under the assumption of the $AR(1)$ input with zero-mean, unit variance and inter-sample autocorrelation coefficient $\rho = 0.95$, the coding gain of the 8-point DCT, 8-point KLT, 16-point DCT and 16-point KLT are $8.8259\,dB$, $8.8462\,dB$, $9.4555\,dB$ and $9.4781\,dB$, respectively.

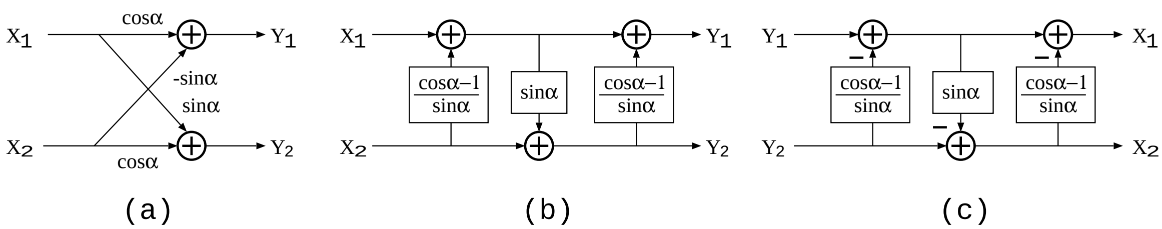


Figure 1. Representation of a rotation by three lifting steps: (a) Givens rotation; (b) Lifting steps; (c) Inverse transform.

2.2. Mean square error (MSE)

To maintain the compatability between the binDCT and the true DCT outputs, the MSE between the DCT and the binDCT coefficients should be minimized. With reasonable assumption of the input signal, the MSE can be explicitly calculated as follows.[27]

Assume $\mathbf{C}$ is the true M-point DCT matrix, and $\mathbf{C}'$ is its approximation, then for a given input column vector $\mathbf{x}$, the error between the 1D M-point DCT coefficients and the approximated transform coefficients is:

$$\mathbf{e} = \mathbf{C}\,\mathbf{x} - \mathbf{C}'\,\mathbf{x} = (\mathbf{C} - \mathbf{C}')\,\mathbf{x} \triangleq \mathbf{D}\,\mathbf{x}, \tag{2}$$

From this the MSE of each transform coefficient can be given by:

$$\epsilon \triangleq \frac{1}{M} E[\mathbf{e^T}\,\mathbf{e}] = \frac{\mathbf{1}}{\mathbf{M}}\,\mathbf{E}[\mathbf{x^T D^T D x}] = \frac{\mathbf{1}}{\mathbf{M}}\,\mathbf{E}[\text{Trace}\{\mathbf{D x x^T D^T}\}] = \frac{\mathbf{1}}{\mathbf{M}}\,\text{Trace}\{\mathbf{D R_{xx} D^T}\}, \tag{3}$$

where $\mathbf{R_{xx}} \triangleq \mathbf{E}[\mathbf{x x^T}]$ is the autocorrelation matrix of the input signal. Hence if we assume the input signal is an $AR(1)$ process with zero-mean, unit variance, and autocorrelation coefficient $\rho = 0.95$, the matrix $\mathbf{R_{xx}}$ can be easily calculated explicitly, thus the MSE can be evaluated by this formula.

Besides the coding gain and the MSE, the stop-band attenuation and the DC leakage will also serve as performance measures in this paper.[26]

3. THE LIFTING SCHEME AND THE GIVENS ROTATION

Fig. 1 shows the decomposition of a Givens rotation by three lifting steps (assuming $\alpha \neq 0$), as suggested by Daubechies and Sweldens.[19] This can be written in matrix form as:

$$\begin{bmatrix} cos\alpha & -sin\alpha \\ sin\alpha & cos\alpha \end{bmatrix} = \begin{bmatrix} 1 & \frac{cos\alpha-1}{sin\alpha} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sin\alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{cos\alpha-1}{sin\alpha} \\ 0 & 1 \end{bmatrix}. \tag{4}$$

Each lifting step is a biorthogonal transform, and its inverse also has a simple lifting structure. That is:

$$\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -x \\ 0 & 1 \end{bmatrix}. \tag{5}$$

As a result, the inverse of the Givens rotation can be represented by lifting steps as:

$$\begin{bmatrix} cos\alpha & -sin\alpha \\ sin\alpha & cos\alpha \end{bmatrix}^{-1} = \begin{bmatrix} cos\alpha & sin\alpha \\ -sin\alpha & cos\alpha \end{bmatrix} = \begin{bmatrix} 1 & -\frac{cos\alpha-1}{sin\alpha} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -sin\alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{cos\alpha-1}{sin\alpha} \\ 0 & 1 \end{bmatrix}. \tag{6}$$

The flow graph of the inverse is given in Fig. 1(c). It can be seen that to inverse a lifting step, we simply need to subtract out what was added at the forward transform. The significance of the lifting structure is that the original signal can still be perfectly reconstructed even if the true lifting parameters are approximated by hardware-friendly dyadic values in both the analysis and the synthesis sides of the transform. This property is the foundation of the proposed binDCT.

Since perfect reconstruction is guaranteed by the lifting structure itself, the remaining problem is to design the lifting parameters such that the binDCT could have similar frequency response and coding gain to the true DCT. Beside, the MSE between the binDCT and the true DCT coefficients should be minimized to maintain their compatability.

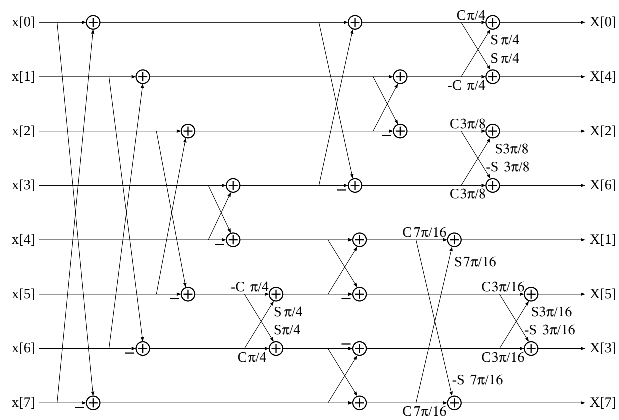


Figure 2. Signal flow graph of Chen's factorization of the 8-point DCT.

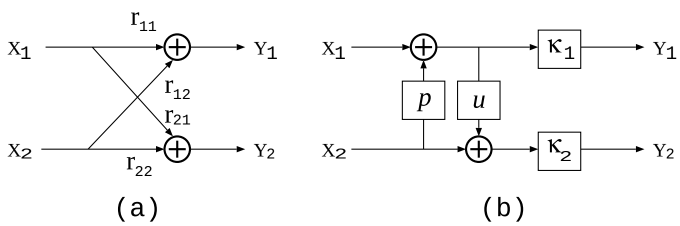


Figure 3. (a) The Givens rotation; (b) The scaled lifting structure.

4. THE SCALED LIFTING STRUCTURE

4.1. Structure

Fig. 2 shows the signal flow graph of Chen's factorization[12,2] of the 1D 8-point DCT matrix. It contains a series of butterflies and five rotation angles. If the two rotations of $\frac{\pi}{4}$ are implemented as a butterfly followed by two multiplications, and each of the other three rotation angles are calculated using 3 multiplications and 3 additions,[13] the overall complexity of the Chen's factorization would be 13 multiplications and 29 additions. Note that a scaling factor of 1/2 should be applied at the end to obtain true DCT coefficients.

The representation of the DCT by butterflies and rotations makes it very suitable for the application of the lifting scheme. Generally, a straightforward substitution of each Givens rotation by three liftings would lead to a DCT with lifting structure. However, it is not the optimal solution in term of the simplicity.

Note that four of the five rotation angles in Fig. 2 are at the end of the transform. This makes it possible to further reduce the complexity by adjusting the aforementioned lifting structure and combining some of the operations in the quantization step.

In Fig. 3, we propose a new structure to represent a Givens rotation by 2 lifting steps and 2 scaling factors. Although it has one more parameter than the traditional structure as shown in Fig. 1, the two scaling factors can be absorbed by the quantization step in implementations. Therefore only 2 lifting steps are left in the transform, which makes it more efficient than the traditional representation. Because of the analogy of this idea to that of the scaled DCT,[3,5,9] we call it the *Scaled Lifting Structure.*

The analytical values of the parameters in the scaled lifting structure can be derived as follows. From the flow graphs in Fig. 3(a), we can obtain the following relationship:

$$\begin{aligned} Y_1 &= r_{11} X_1 + r_{12} X_2, \\ Y_2 &= r_{21} X_1 + r_{22} X_2. \end{aligned} \tag{7}$$

Similarly, the outputs of the scaled lifting structure as given in Fig. 3(b) can be written as:

$$\begin{aligned} Y_1 &= \kappa_1 \left(X_1 + p\, X_2\right) = \kappa_1 X_1 + \kappa_1\, p\, X_2, \\ Y_2 &= \kappa_2 \left(u\left(X_1 + p\, X_2\right) + X_2\right) = \kappa_2\, u\, X_1 + \kappa_2 \left(1 + p\, u\right) X_2. \end{aligned} \tag{8}$$

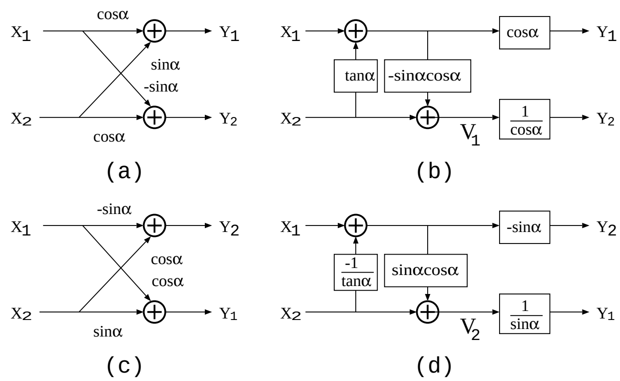


Figure 4. (a) The Givens rotation for $\frac{3\pi}{8}$, $\frac{7\pi}{16}$ and $\frac{3\pi}{16}$ in Fig. 2; (b) The solution of the scaled lifting structure; (c) The flipped Givens rotation; (d) Solution of the flipped and scaled lifting structure.

By equalizing the coefficients of X_1 and X_2 in Eq. 7 and Eq. 8, the four lifting parameters can be obtained as:

$$
\begin{aligned}
p &= \frac{r_{12}}{r_{11}}, \\
u &= \frac{r_{11}\,r_{21}}{r_{11}\,r_{22} - r_{21}\,r_{12}}, \\
\kappa_1 &= r_{11}, \\
\kappa_1 &= \frac{r_{11}\,r_{22} - r_{21}\,r_{12}}{r_{11}}.
\end{aligned} \tag{9}
$$

These analytical solutions are the start points in obtaining binDCTs with different complexities and performances.

4.2. Dyadic Approximations

The property of the listing structure allows us to adjust the lifting parameters without losing the perfect reconstruction of the signals. Therefore, from the analytical expressions as given in Eq. 4 and 9 we can obtain their dyadic approximations by proper truncations, which enables fast implementations with only shift and addition operations.

For example, one of the scaled lifting parameters of the rotation angle $\frac{3\pi}{16}$ in Fig. 2 is $u = 0.461939766\ldots$, which in binary expression is $0.011101\ldots$. This can be approximated as $1/2$, $7/16$ or $15/32$, with different accuracies. Hence the analytical values make it easy to accomplish different tradeoffs between the performance and the complexity of the fast transform, according to the requirements of different applications.

Note that in the implementation of certain dyadic values, simple equivalent relationships can be exploited to minimize the cost. For example, $\frac{7}{16}x$ should be implemented as $\frac{1}{2}x - \frac{1}{16}x$ instead of $\frac{1}{4}x + \frac{1}{8}x + \frac{1}{16}x$. The same rule can be used to other values like $\frac{11}{16}$ and $\frac{15}{32}$.

4.3. Permuted structure and sensitivity analysis

In this part, we will analyze the effect of the truncation error on the performance of the scaled lifting structure, and a permuted version of the scaled lifting structure will be proposed to improve its performance in certain circumstances.

In Fig. 4(a), we redraw the general format of the rotation angles $\frac{3\pi}{8}$, $\frac{7\pi}{16}$ and $\frac{3\pi}{16}$ in Fig. 2. The solutions of its corresponding scaled lifting structure can be obtained by Eq. 9, as shown in Fig. 4(b).

The signal V_1 in Fig. 4(b) can be expressed as:

$$V_1 = -sin\alpha\, cos\alpha\, X_1 + cos^2\alpha\, X_2, \tag{10}$$

Eq. 10 shows that for Givens rotations as shown in Fig. 4(a), the coefficient $cos^2\alpha$ in V_1 would be very close to 0 if the rotation angle is close to $\frac{\pi}{2}$. Therefore large relative error would be resulted when p and u are truncated or rounded, which would lead to drastic change in the frequency response of the generated binDCT.

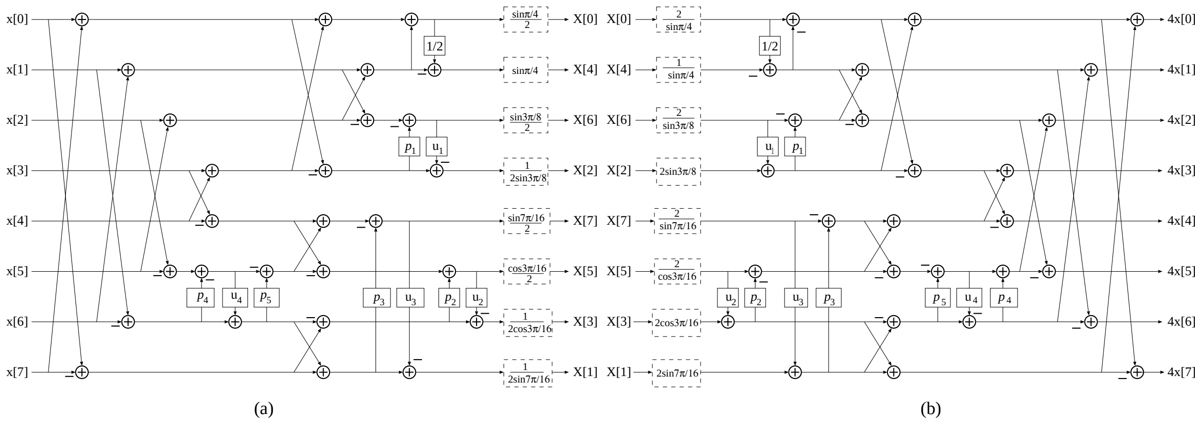


Figure 5. General structure of binDCT family from Chen's factorization: (a) Forward transform; (b) Inverse transform.

Another problem is that the lifting parameter $tan\alpha$ would be much greater than 1 when the angle α is close to $\pi/2$. This increases the dynamic range of the intermediate results, and is not desired in both software and hardware implementations.

However, in this case, a simple permutation of the output as shown in Fig. 4(c) will lead to a much more robust scaled lifting structure, where the positions of the two output signals are switched. Since the rotation coefficients are changed accordingly, the new transform is equivalent to the previous one. The general expression in Eq. 9 is still valid for this case, and the corresponding scaled lifting parameters are given in Fig. 4(d).

The signal V_2 in Fig. 4(d) is now given by:

$$V_2 = sin\alpha\, cos\alpha\, X_1 + sin^2\alpha\, X_2. \tag{11}$$

Thus the coefficient of X_2 in signal V_2 changes from $cos^2\alpha$ to $sin^2\alpha$, which is more robust to truncation errors for rotation angles close to $\pi/2$. Besides, the augment of the dynamic range in Fig. 4(b) is also avoided in the new version, as the first lifting parameter becomes $\frac{1}{tan\alpha}$ now, which is less than 1 for rotation angles close to $\pi/2$.

This analysis reveals that flip is necessary for $\frac{3\pi}{8}$ and $\frac{7\pi}{16}$ in Chen's factorization of the DCT matrix, since $cos^2(\frac{3\pi}{8}) = 0.14645$, and $cos^2(\frac{7\pi}{16}) = 0.03806$, both are too small to be accurately represented by p and u with finite length. However, it is safe to keep the output order in the rotation of $\frac{3\pi}{16}$.

5. BINDCT FAMILY FROM CHEN'S FACTORIZATION

5.1. General Structure

From the above analysis, we can obtain the general structure of the forward and inverse binDCT from Chen's factorization, as shown in Fig. 5, where the intermediate rotation angle of $\pi/4$ is implemented as the traditional 3-lifting structure, and the permuted version of the scaled lifting structure is used for $\frac{3\pi}{8}$ and $\frac{7\pi}{16}$.

The rotation of $\frac{\pi}{4}$ between $X[0]$ and $X[4]$ is also implemented by the scaled lifting structure, instead of a scaled butterfly. The purpose of this operation is to make all subbands experience the same number of butterflies during the forward and inverse transforms. Since the multiplication of two butterflies introduces a scaling factor of 2, the combination of the forward and inverse transforms thus generates a uniform scaling factor of 4 for all subbands, which becomes 16 for 2D transform. This can be compensated by a simple shift operation.

The scaling factors in the dash boxes will be absorbed by the quantization step. They can also be bypassed if both the encoder and decoder use the same transform. Note that some sign manipulations are involved here to make all the scaling factors positive.

Table 1 lists the analytical values and some possible dyadic values for all the lifting parameters in Fig. 5. These options are obtained by truncating or rounding the corresponding analytical values with different accuracies. In Table 2, we present some configurations of this binDCT family, and Fig. 6 presents the frequency responses of the

Table 1. Some possible dyadic values for lifting parameters in Fig. 5.

Angle	Symbol	Analytical Value	Binary	Option 1	Option 2	Option 3
$\frac{3\pi}{8}$	p_1	0.41421356237310	0.011010 ...	3/8	7/16	13/32
	u_1	0.35355339059327	0.010110 ...	5/16	3/8	11/32
$\frac{3\pi}{16}$	p_2	0.66817863791930	0.101010 ...	7/8	5/8	11/16
	u_2	0.46193976625564	0.011101 ...	1/2	7/16	15/32
$\frac{7\pi}{16}$	p_3	0.19891236737966	0.001100 ...	1/4	3/16	13/64
	u_3	0.19134171618254	0.001100 ...	1/4	3/16	13/64
$\frac{\pi}{4}$	p_4	0.41421356237310	0.011010 ...	3/8	7/16	13/32
	u_4	0.70710678118655	0.101101 ...	5/8	3/4	11/16
	p_5	0.41421356237310	0.011010 ...	3/8	7/16	13/32

Table 2. Some configurations of binDCT from Chen's factorization.

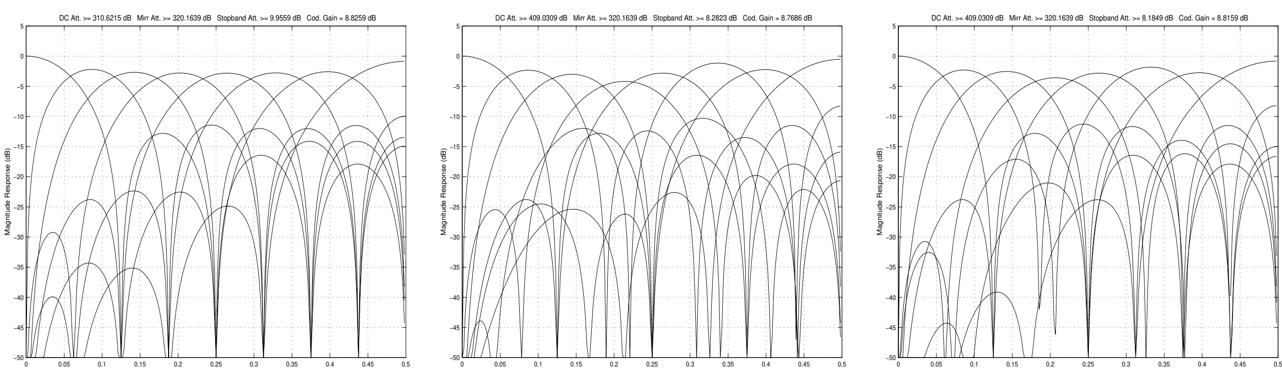
Config.	p_1	u_1	p_2	u_2	p_3	u_3	p_4	u_4	p_5	Shifts	Adds	MSE	C_g (dB)
1	1/2	1/2	1	1/2	1/4	1/4	1/2	3/4	1/2	9	28	$2.30E-3$	8.7686
2	1/2	3/8	7/8	1/2	3/16	1/4	7/16	3/4	3/8	14	33	$5.78E-4$	8.8033
3	3/8	3/8	7/8	1/2	3/16	3/16	7/16	11/16	3/8	17	36	$4.19E-4$	8.8159
4	7/16	3/8	5/8	7/16	3/16	3/16	7/16	11/16	3/8	19	37	$8.49E-5$	8.8220
5	13/32	11/32	11/16	15/32	3/16	3/16	7/16	11/16	3/8	21	40	$3.37E-5$	8.8233
6	7/16	3/8	5/8	7/16	3/16	3/16	13/32	11/16	13/32	21	39	$5.70E-5$	8.8240
7	13/32	11/32	11/16	15/32	3/16	3/16	13/32	11/16	13/32	23	42	$1.11E-5$	8.8251

true DCT and some binDCT examples, where the configuration 3 is exactly the same as our previously proposed binDCT version A.[21–23]

The complexities of the configurations in Table 2 range from 9 shifts and 28 additions to 23 shifts and 42 additions. The configuration with 23 shifts has a coding gain of $8.8251\,dB$, which is almost the same as the $8.8259\,dB$ of the true DCT. Even the 9-shift version has a good coding gain of $8.7686\,dB$. However, it has worse MSE than other finer versions. Note that in measuring the MSE according to Eq. 3, we use the floating-point values of the scaling factors, which are always rounded to integers in actual implementations. Therefore the actual MSE would be slightly different from the theoretical ones.

5.2. Performance comparison between the two types of scaled lifting structures

In this part, we use the subband-7 of the binDCT according to Config. 3 of Table 2 to demonstrate the difference of the two types of scaled lifting structures. In Fig. 7(a), the frequency response of the binDCT is obtained when the



(a) (b) (c)

Figure 6. Frequency responses of (a) The true DCT; (b) Config. 1 of Table 2: 9 Shifts and 28 Adds; (c) Config. 3: 17 shifts and 36 Adds.

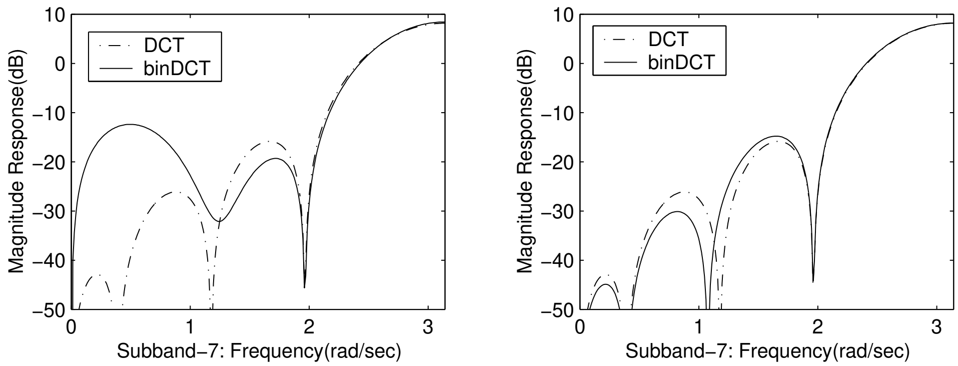


Figure 7. Frequency response of subband 7 in Chen's factorization-based binDCT: (a) $X[1]$ and $X[7]$ are not permuted; (b) $X[1]$ and $X[7]$ are permuted as in Fig. 5.

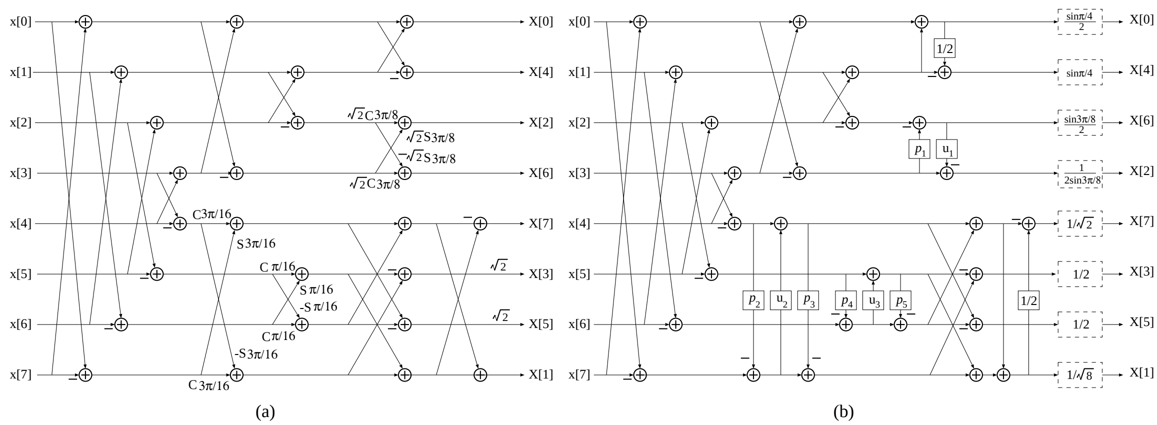


Figure 8. (a) Loeffler's factorization of the 8-point DCT; (b) binDCT from Loeffler's factorization.

angle $7\pi/16$ is implemented as the normal scaled lifting structure. The analytical values of the lifting parameters are $p = 5.027339492$ and $u = -0.19134172$, and they are approximated as $5\frac{3}{128} = 5.0234375$ and $-3/16 = -0.1875$. The result in Fig. 7(b) is obtained according to the Config. 3, i.e., the output $X[7]$ and $X[1]$ are permuted, as in Fig. 5.

As shown in Fig. 7, for this kind of rotation angle, the frequency response of the output is distorted dramatically if the outputs are not permuted, even though the lifting parameters p and u approximate their analytical values quite well. On the contrary, the frequency response of the permuted version agrees very well with the true DCT, and therefore has higher coding gain and smaller MSE.

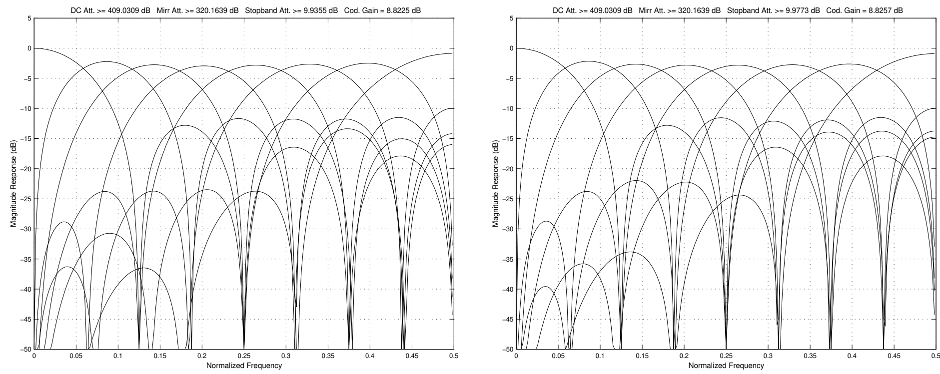
6. LOEFFLER'S FACTORIZATION-BASED BINDCT

6.1. 8-point binDCT family

The aforementioned design method for the binDCT can also be applied to other factorizations of the DCT matrix, provided that all the intermediate multiplications in the factorization are in the format of the Givens rotations. Besides, it is also applicable to other DCT sizes, such as the 16-point DCT. In this section, we will give more results with the Loeffler's factorizations.[13]

Fig. 8(a) shows the Loeffler's factorization of the 8-point DCT matrix, which only needs 11 multiplications and 29 additions. This achieves the lower bound for 8-point DCT.[16,17] One of its variations is adopted by the Independent JPEG Group's implementations of the JPEG standard.[28] Note that this factorization requires a uniform scaling factor of $1/\sqrt{8}$ at the end of the flow graph to obtain true DCT coefficients. In the 2D transform, this becomes 1/8, which can be easily implemented by a shift operation.

Fig. 8(b) shows the structure of the corresponding binDCT. The first four subbands are exactly the same as the binDCT from Chen's factorization. Since the other two rotation angles are not at the end of the flow graph,



(a) (b)

Figure 9. Frequency responses of binDCTs from Loeffler's factorization: (a) 16 Shifts and 34 Adds; (b) 22 shifts and 40 Adds.

Table 3. Some possible dyadic values for lifting parameters in the binDCT from Loeffler's factorization.

Angle	Symbol	Analytical Value	Binary	Option 1	Option 2	Option 3
$\frac{3\pi}{8}$	p_1	0.414213562	0.011010 ...	3/8	7/16	13/32
	u_1	0.353553391	0.010110 ...	5/16	3/8	11/32
$\frac{3\pi}{16}$	p_2	0.303346683	0.010011 ...	1/4	5/16	19/32
	u_2	0.555570233	0.100011 ...	1/2	9/16	35/64
	p_3	0.303346683	0.010011 ...	1/4	5/16	19/32
$\frac{\pi}{16}$	p_4	0.098491403	0.000110 ...	1/16	1/8	3/32
	u_3	0.195090322	0.001100 ...	1/8	1/4	3/16
	p_5	0.098491403	0.000110 ...	1/16	1/8	3/32

we represent them with the standard 3 lifting steps. Note that the final butterfly to obtain $X[7]$ and $X[1]$ is also implemented as 2 liftings to maintain the same number of butterflies for each subband, which leads to a uniform scaling factor after inverse binDCT transform.

The analytical values of the lifting parameters in Fig. 8(b) can be easily calculated, and the results are summarized in Table 3, together with some possible dyadic approximations. Various binDCT configurations can be obtained through this table, and some examples are given in Table 4. The frequency responses of some configurations are presented in Fig. 9. As shown, the new type has a slightly better performance than the previous one in term of the coding gain and MSE that can be obtained within a given complexity. This is due to the reduced number of the Givens rotations in the Loeffler's factorization. For example, the configuration with 22 shifts can achieve a stunning coding gain of 8.8257 dB and a MSE of $8.18E-6$.

6.2. 16-point binDCT family

A Givens rotation-based factorization of the 16-point DCT is also proposed by Loeffler *et. al.*,[13] which needs 31 multiplications and 81 additions. Although the lower bound for the number of multiplications of 16-point DCT is 26,[16] the Loeffler's 16-point factorization is so far one of the most efficient solutions.

Table 4. Family of 8-point binDCTs from Loeffler's factorization.

Config.	p_1	u_1	p_2	u_2	p_3	p_4	u_3	p_5	Shifts	Adds	MSE	C_g (dB)
1	1/2	1/2	1/4	1/2	1/4	1/8	1/4	1/8	10	28	$6.92E-4$	8.7716
2	3/8	1/2	1/4	1/2	1/4	1/8	3/16	3/32	13	31	$3.58E-4$	8.8027
3	7/16	3/8	1/4	9/16	5/16	1/8	3/16	3/32	16	34	$4.01E-5$	8.8225
4	13/32	11/32	5/16	9/16	5/16	3/32	3/16	3/32	20	38	$1.12E-5$	8.8242
5	13/32	11/32	19/64	9/16	19/64	3/32	3/16	3/32	22	40	$8.18E-6$	8.8257

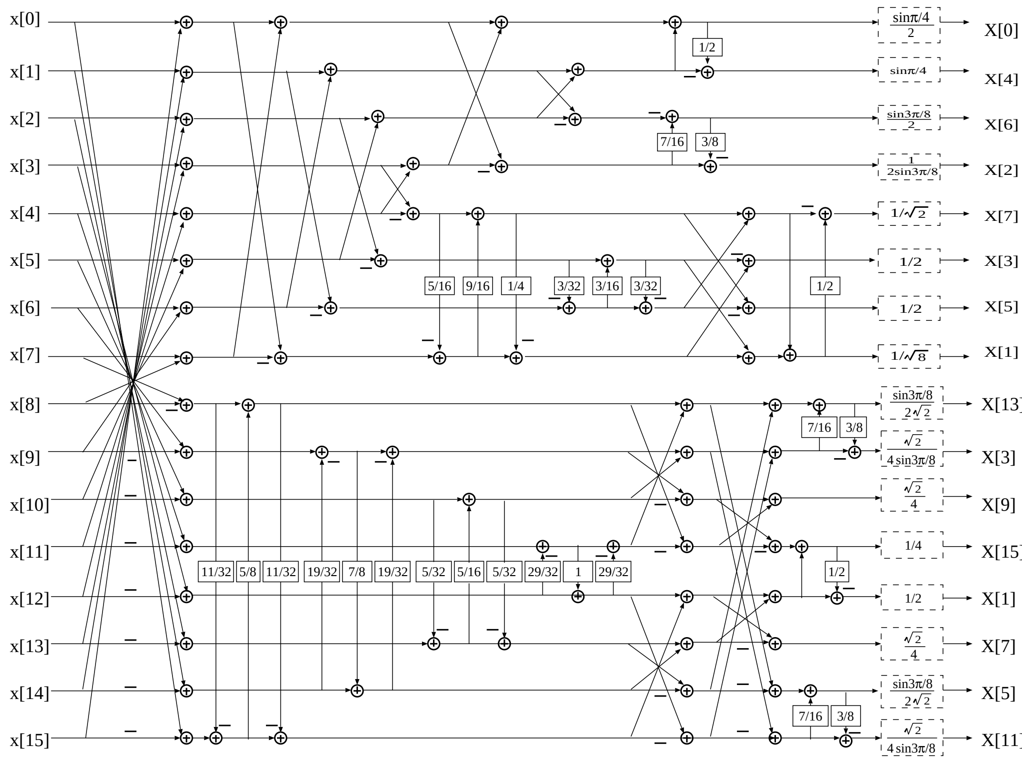
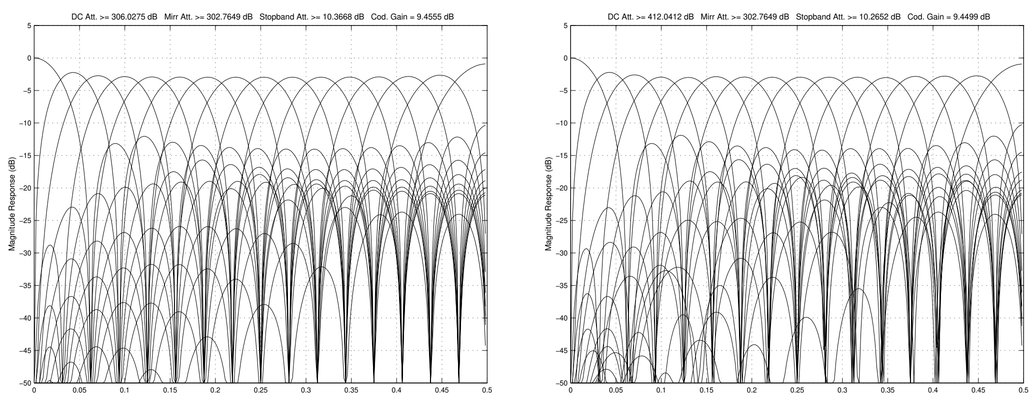


Figure 10. A 16-point binDCT from Loeffler's factorization: 51 shifts and 106 Adds.



(a) (b)

Figure 11. (a) Frequency response of the 16-point DCT; (b) Frequency response of the 16-point binDCT as shown in Fig. 10: MSE: 8.4952E-5.

With our proposed design method, a family of 16-point binDCT can be easily obtained from this factorization. The general structure and an example is given in Fig. 10. As shown, the even part of the 16-point DCT is exactly the same as the 8-point DCT. The example in Fig. 10 requires 51 shifts and 106 additions. Its frequency response is given in Fig. 11, together with that of the true 16-point DCT. The coding gain of this binDCT is $9.4499\,dB$, which is very close to the $9.4555\,dB$ of the true 16-point DCT. The MSE of this approximation is $8.4952E-5$.

7. EXPERIMENTAL RESULTS

The proposed binDCT families have been implemented according to the framework of the JPEG standard, based on the source code from the Independent JPEG Group (IJG).[28] We replace the DCT part by the proposed binDCT,

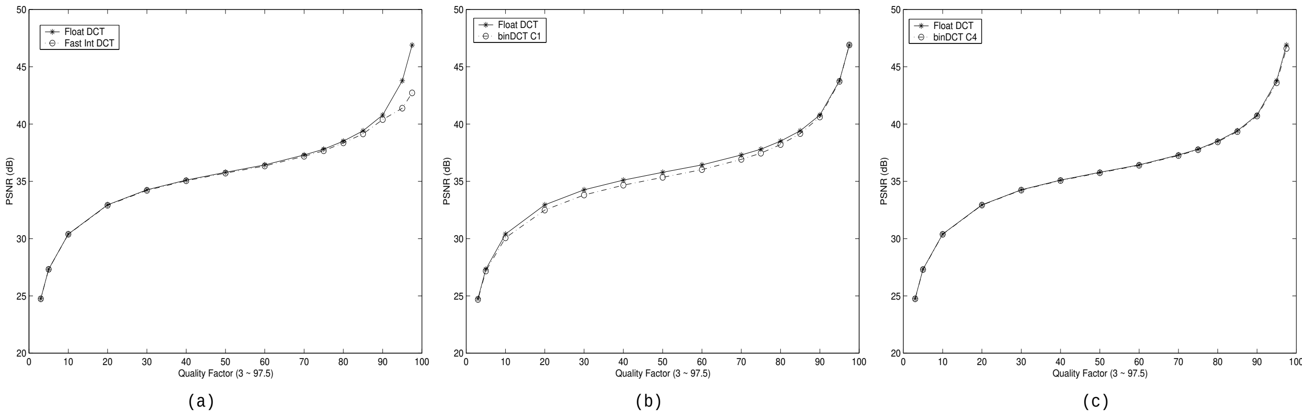


Figure 12. PSNRs of (a) Floating DCT and fast integer DCT; (b) Floating DCT and Config. 1 of the binDCT from Chen's factorization; (c) Floating DCT and Config. 4 of the binDCT from Chen's factorization.

and the quantization matrix is modified to incorporate the binDCT scaling factors. Fig. 12 compares the PSNR results of the reconstructed Lena image with two configurations of the binDCT, the floating version and the fast integer version of the IJG's code.[9,3] The Config. 1 and 4 of the binDCT family from Chen's factorization are chosen in this example. In obtaining these results, the same kind of transform is used in both the forward and inverse DCT.

It can be seen from Fig. 12 that the performances of both examples of the binDCT are very close to the floating DCT implementation. In particular, when the quality factor is below 90, the difference between the Config. 4 of this type of binDCT and the floating DCT is less than $0.1dB$, which is negligible. The performance degradation of the 9-shift and 28-addition version of the binDCT is also below $0.5dB$. Besides, when the quality factor is above 90, the performance of the proposed binDCT is much better than the fast integer version of the IJG's code.

8. CONCLUSION

A systematic approach to design families of fast multiplierless approximations of the DCT with the lifting scheme is presented, based on factorizations of the DCT matrix with Givens rotations. A scaled lifting structure is proposed to reduce the complexity of the transform, and its analytical solution is derived. Different dyadic values can be obtained through proper finite-length approximations of the analytical solutions. Besides, a sensitivity analysis is developed for the scaled lifting structure, which shows that for certain rotation angles, a permuted version is more robust to truncation errors than the normal version.

Various approximations of the 8-point and 16-point DCT with different complexities are presented, based on Chen's and Loeffler's factorizations, respectively. These results are very close to the floating DCT in terms of coding gains, frequency responses, and mean square errors of DCT coefficients. These transforms enable fast and low-cost software or hardware implementations with only shift and addition operations, making them very suitable for real-time and mobile computing applications.

REFERENCES

1. N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Computer* **Vol. C-23**, pp. 90–93, Jan., 1974.
2. K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, New York, 1990.
3. W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, New York, 1993.
4. J. Mitchhell, W. Pennebaker, C. E. Fogg, and D. J. LeGall, *MPEG video compression standard*, Chapman and Hall, New York, 1997.
5. V. Bhaskaran and K. Konstantinides, *Image and video compression standards: algorithms and architectures*, Kluwer Academic Publishers, Boston, 1997.

6. S. C. Chan and K. L. Ho, “A new two-dimentional fast cosine transform algorithm,” *IEEE Trans. Signal Processing* **Vol. 39, No. 2**, pp. 481–485, Feb., 1991.
7. Y. Jeong, I. Lee, T. Yun, G. Park, and K. Park, “A fast algorithm suitable for dct implementation with integer multiplication,” *Proceedings of 1996 IEEE TENCON - Diginal Signal Processing Applications* , pp. 784–787, 1996.
8. J. Makhoul, “A fast cosine transform in one and two dimentions,” *IEEE Trans. Acoustics, Speech, and Signal Processing* **Vol. ASSP-28, No. 1**, pp. 27–34, Feb., 1980.
9. Y. Arai, T. Agui, and M. Nakajima, “A fast dct-sq scheme for images,” *Trans. IEICE* **E-71(11)**, p. 1095, 1988.
10. D. Hein and N. Ahmed, “On a real-time walsh-hadamard cosine transform image processor,” *IEEE Trans. Electromag. Compat.* **Vol. EMC-20**, pp. 453–457, Aug. 1978.
11. H. Malvar, “Fast computation of the discrete cosine transform and the discrete hartley transform,” *IEEE Trans. Acoustics, Speech, and Signal Processing* **Vol. ASSP-35**, pp. 1484–1485, Oct. 1987.
12. W. Chen, C. Harrison, and S. Fralick, “A fast computational algorithm for the discrete cosine transform,” *IEEE Trans. Communications* **Vol. COM-25, No. 9**, pp. 1004–1011, 1977.
13. C. Loeffler, A. Lightenberg, and G. Moschytz, “Practical fast 1d dct algorithms with 11 multiplications,” *Proc. IEEE ICASSP* **Vol. 2**, pp. 988–991, Feb. 1989.
14. B. G. Lee, “A new algoithm to compute the discrete cosine transform,” *IEEE Trans. Acoustics, Speech, and Signal Processing* **Vol. ASSP-32**, pp. 1243–1245, Dec. 1984.
15. H. S. Hou, “A fast recursive algorithm for computing the discrete cosine transform,” *IEEE Trans. Acoustics, Speech, and Signal Processing* **Vol. ASSP-35**, pp. 1455–1461, Oct. 1987.
16. P. Duhamel and H. H’Mida, “New 2^n dct algorithms suitable for vlsi implementation,” *Proc. IEEE ICASSP* , pp. 1805–1808, 1987.
17. E. Feig and S. Winograd, “On the multiplicative complexity of discrete cosine transform,” *IEEE Trans. Information Theory* **Vol. 38**, pp. 1387–1391, Jul. 1992.
18. W. Sweldens, “The lifting scheme: A custom-design construction of biorthogonal wavelets,” *Appl. Comput. Harmon. Anal* **Vol. 3, No. 2**, pp. 186–200, 1996.
19. I. Daubechies and W. Sweldens, “Factoring wavelet transforms into lifting step,” *Journal Fourier Anal. Appl.* **Vol. 4, No. 3**, pp. 247–269, 1998.
20. P. Vaidyanathan and P. Hoang, “Lattice structures for optimal design and robust implementation of two-abdn perfect reconstruction qmf banks,” *IEEE Trans. Accoust. Speech and Signal Process.* **Vol. 36**, pp. 81–94, 1988.
21. T. Tran, “Fast multiplierless approximation of the dct,” *Proc. of the 33rd Annual Conf. on Information Sciences and Systems* , pp. 933–938, Mar. 1999.
22. T. Tran, “A fast multiplierless block transform for image and video compression,” *Proc. of the IEEE ICIP* **Vol. 3**, pp. 822–826, Oct. 1999.
23. T. Tran, “The bindct: fast multiplierless approximation of the dct,” *IEEE Signal Processing Letters* **Vol. 7, No. 6**, pp. 141–145, Jun. 2000.
24. N. Jayant and P. Noll, *Digital coding of waveforms*, Prentice-Hall, New Jersey, 1984.
25. J. Katto and Y. Yasuda, “Performance evaluation of subband coding and optimization of its filter coefficients,” *SPIE Proc. Visual Commun. Image Process.* , pp. 95–106, Boston, MA, Nov. 1991.
26. T. Tran, R. L. de Queiroz, and T. Nguyen, “Linear-phase perfect reconstruction filter bank: Lattice strucutre, design, and applications in image coding,” *IEEE Trans. Signal Processing* **Vol. 48, No. 1**, pp. 133–147, Jan. 2000.
27. R. Queiroz, “On unitary transform approximations,” *IEEE Signal Processing Letters* **Vol. 5, No. 2**, pp. 46–47, Feb. 1998.
28. *ftp://ftp.uu.net/graphics/jpeg* .