OPTIMAL UNIFIED IIR ARCHITECTURES FOR TIME-RECURSIVE DISCRETE SINUSOIDAL TRANSFORMS †

K.J.R. Liu, C.T. Chiu, R. K. Kolagotla‡ and J .F. JáJá
Electrical Engineering Department and Institute of Systems Research,
University of Maryland, College Park, MD 20742 USA

ABSTRACT

An optimal unified architecture that can efficiently compute the Discrete Cosine, Sine, Hartley, Fourier, Lapped Orthogonal, and Complex Lapped transforms for a continuous input data stream is proposed. This structure uses only half as many multipliers as the previous best known scheme [1]. The proposed architecture is regular, modular, and locally-connected. In the realization of the DCT, only $2N-2$ multipliers are needed. We provide a theoretical justification by showing that any discrete transform whose basis functions satisfy the *Fundamental Recurrence Formula* has a second-order autoregressive structure in its filter realization. We also demonstrate that dual generation transform pairs share the same autoregressive structure.

1. INTRODUCTION

Discrete sinusoidal transforms play an important role in various digital signal processing applications, such as spectrum analysis, speech and image signal processing, computer tomography, data compression and reconstruction, *etc.* The development of efficient algorithms and architectures for discrete sinusoidal transforms that are suitable for hardware implementation has been an interesting topic for decades. A parallel lattice structure that can dually generate the DCT and DST simultaneously was proposed by Liu and Chiu [1]. The structure is regular, modular, locally-connected and suitable for VLSI implementation. The number of multipliers used in that lattice structure is $O(N)(N$ is the transform size), in contrast to $O(N\log N)$, which is common in many other fast algorithms with a buffered data, irregular, and globally connected scheme [2,3]. For example, for the 1-D DCT case, the lattice structure needs $6N-8$ multipliers [1], while $((N/2)\log N)$ multipliers are used in reference [2,3]. The lattice structure is superior than other computing schemes except possibly when N is very small. In this paper, we propose an optimal unified IIR architecture for the discrete sinusoidal transforms, which preserves the advantages of the lattice architecture, while reducing the hardware complexity in half. It is optimal in the sense that the number of multipliers used is minimum, and both speed and area are asymptotically optimal. A theoretical basis that allows us to show that all the resulting unified filter architectures have a similar second-order autoregressive structure is provided in Section 3.

†The work is partially supported by the NSF grant ECD-8803012.

‡Now with IBM Corp. Burlington, VT05452.

2. OPTIMAL TIME-RECURSIVE ARCHITECTURES

2.1 Optimal Unified IIR Structures

Input data arrive serially in most real-time signal processing applications. Our approach is to view the lattice modules in [1] as linear time invariant systems which transform input sequences into output sequences. The transfer functions can be determined by the time difference equations obtained from the lattice modules. For all these discrete sinusoidal transforms, the transfer functions for the k-th transformed component have the following form

$$H(z) = ((-1)^m - z^{-n})\,\frac{(N_1 + N_2 z^{-1})}{(1 - D_1 z^{-1} + D_2 z^{-2})}. \tag{1}$$

The coefficients, m, n, N_1, N_2, D_1, and D_2, for different discrete sinusoidal transforms are listed in Table 1. Note that $C(k)$ in Table 1 is the the normalization constant for the k-th transformed component. From the transfer functions derived above, we observe that these transforms can be realized by

0-7803-0946-4/93 $3.00

	m, n	D_1, D_2	N_1	N_2
DCT	k, N	$2\cos\left(\frac{\pi k}{N}\right)$, 1	$C(k)\sqrt{\frac{2}{N}}$ $\cos\left(\frac{\pi k}{2N}\right)$	$-C(k)\sqrt{\frac{2}{N}}$ $\cos\left(\frac{\pi k}{2N}\right)$
DST	k, N	$2\cos\left(\frac{\pi k}{N}\right)$, 1	$-C(k)\sqrt{\frac{2}{N}}$ $\sin\left(\frac{\pi k}{2N}\right)$	$-C(k)\sqrt{\frac{2}{N}}$ $\sin\left(\frac{\pi k}{2N}\right)$
DHT	$0, N$	$2\cos\left(\frac{2\pi k}{N}\right)$, 1	$\frac{1}{\sqrt{N}}\left[\cos\left(\frac{2\pi k}{N}\right) - \sin\left(\frac{2\pi k}{N}\right)\right]$	$\frac{1}{\sqrt{N}}$
DFT	$0, N$	$2\cos\left(\frac{2\pi k}{N}\right)$, 1	$\frac{1}{\sqrt{N}}\left[\cos\left(\frac{2\pi k}{N}\right) + j\sin\left(\frac{2\pi k}{N}\right)\right]$	$\frac{1}{\sqrt{N}}$
CLT	$0, 2N$	$\exp j2\theta_k$ $2\cos\left(\frac{\pi}{2N}\right)$, $\exp j4\theta_k$	$\sin\left(\frac{\pi}{4N}\right)^2$ $\exp j\theta_k$	$\sin\left(\frac{\pi}{4N}\right)$ $\cos\left(\frac{\pi}{4N}\right)$ $(-1)^k$ $\exp j4\theta_k$

Table 1: Coefficients of the universal IIR filter structure for the DXT($^*\theta_k = \frac{(2k+1)\pi}{4N}$).

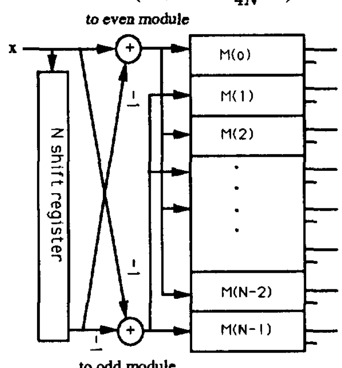

Figure 1: The parallel IIR filter structure for 1-D DXT.

a parallel structure consisting of a shift register array of size N, two adders and an IIR array made up of N DXT modules as shown in Fig. 1. Each IIR module is a second order autoregressive IIR filter as shown in Fig. 2. From Table 1, we further note that the DCT and DST, DFT and DHT share the same denominator and can be simultaneously generated by using an IIR filter structure with three and four multipliers respectively as depicted in Figs. 3 and 4. The transfer function given in Table 1 for the LOT/CLT is in complex form. We will show in the following how to realize the CLT using real operations. The definition of the CLT in [4] can be rewritten as

$$
\begin{aligned}
X_{clt}(k) &= (-1)^k j \frac{1}{\sqrt{N}} \sum_{n=0}^{2N-1} x(n) \sin\frac{(2n+1)\pi}{4N} \cdot \\
&\quad \exp\{-j\frac{(2k+1)(2n+1)\pi}{4N}\}, \\
&\quad k = 0, 1, \ldots, N-1. \qquad (2)
\end{aligned}
$$

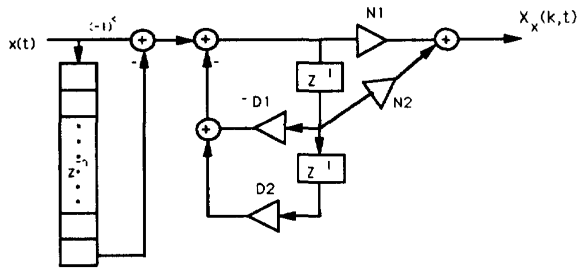

Figure 2: The universal IIR filter module.

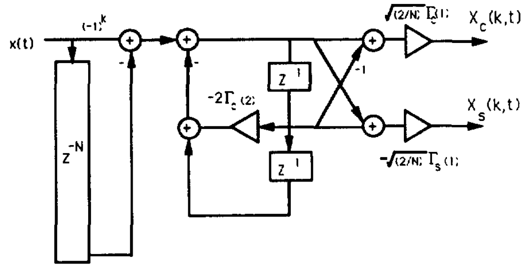

Figure 3: The IIR filter structure for the DCT and DST.

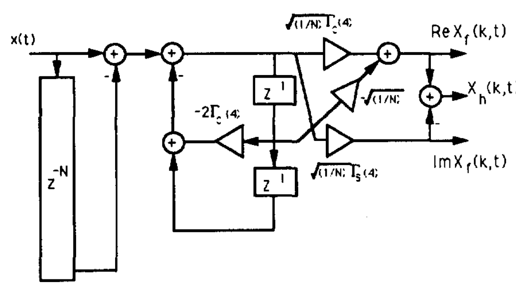

Figure 4: The IIR filter structure for the DHT and DFT.

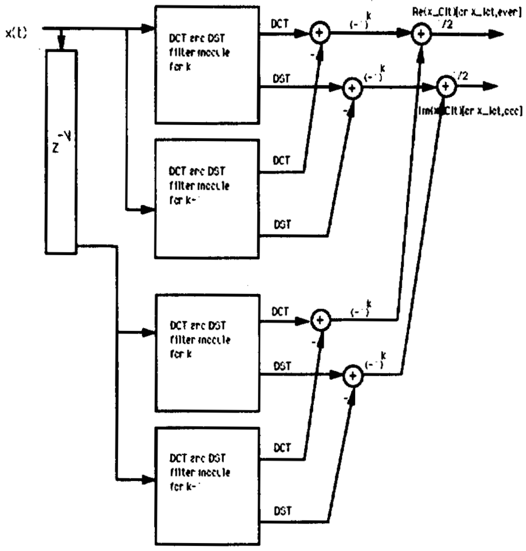

Figure 5: The IIR filter structure for real operation of the LOT and CLT.

If we define another transform with basis functions t_{nk}, then the CLT can be expressed in the form of [4]

$$X_{clt}(k) = \frac{1}{2}(-1)^k \sum_{n=0}^{N-1} x(n)[t_{nk} - t_{n(k+1)}] \quad (3)$$

$$+\frac{1}{2}(-1)^k \sum_{n=N}^{2N-1} x(n)[t_{nk} + t_{n(k+1)}],$$

where $t_{nk} = \frac{1}{N}\exp\frac{j(2n+1)k\pi}{2N}$. This leads to the CLT architecture as shown in Fig. 5, in which the t_{mn} are generated by using the DCT and DST dual generating circuit as depicted in Fig. 3. The numbers of multipliers and adders required for computing individual transforms by using the IIR filter structures are summarized in Table 2. From Table 2, we observe that by using the IIR structure the number of multipliers required for the DCT has been reduced from $6N-8$ to $2N-2$. Table 3 provides a comparison of the optimal unified IIR structure with other well-known algorithms.

2.2 Unified inverse IIR structure

Inverse transforms are important in retrieving original information in digital communication systems. The inverses of the discrete sinusoidal transforms, which have similar transfer formulae are also carried out under the same methodology. The inverse transforms of the discrete sinusoidal transforms are the same as their direct versions except for the appearance of the normalization constants, $C(k)$'s. For example, the transfer function of the inverse DCT(as defined in [1]) is

$$H_{ic}(z) = \sqrt{\frac{2}{N}}\frac{z^{-N-1} - \cos\theta z^{-N} + (-1)^n \sin\theta}{1 - 2\cos\theta z^{-1} + z^{-2}},$$

$$+\sqrt{\frac{2}{N}}(\sqrt{\frac{1}{2}} - 1)z^{-(N-1)}, \quad (4)$$

where $\theta = \frac{\pi(n+.5)}{N}$. If we perform the block transform instead of sliding window transform, then the z^{-N-1} and z^{-N} components in the numerator can be eliminated because of the reset operation. In Fig. 6, we show the optimal unified IIR implementation of the inverse DCT module under block transform. The number of multipliers required for the inverse DCT is $2N-1$. The additional branch of multiplier is shared by the N IIR array with a delay of $N-1$ cycles. The difference in the direct and inverse transform formula can be rectified by adding one additional branch of multipliers to a whole parallel IIR structure and changing the multiplication coefficients.

Transforms	multipliers	adders
DCT	$2N-2$	$3N+2$
DST	$2N-2$	$3N+2$
DHT	$2N$	$3N+1$
DFT	$3N-2$	$3N+1$
LOT*	$4N$	$4N$
CLT*	$4N$	$4N$
DCT and DST	$3N-3$	$4N+2$
DHT and DFT	$3N-2$	$4N+1$

Table 2: Number of multipliers and adders for different transforms with IIR filter realizations(Here * denotes complex operation)

	direct IIR	lattice	Lee[2]	Hou[3]
No. of Multipliers	$2N-2$	$6N-8$	$(N/2)$ *$\log N$	$(N/2)$ *$\log N$
No. of adders	$3N+2$	$5N-1$	$(3N/2)\log N$ $-N+1$	same as Lee
limitation on N	no	no	2^n	2^n
commun.	local	local	global	global
I/O	$SIPO$	$SISO$	$PIPO$	$PIPO$

Table 3: Comparison of different DCT algorithms

3. THEORETICAL BASIS

All the resulting unified filter architectures mentioned above have a similar second-order autoregressive structure with the minimum number of multipliers. A theoretical basis for this fact is provided in

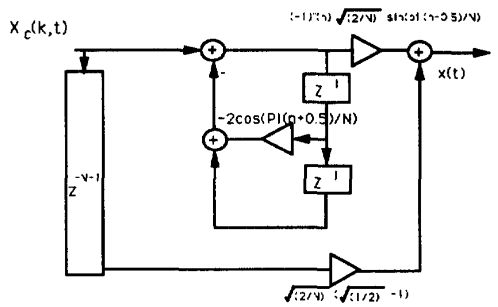

Figure 6: The IIR filter structure for the IDCT.

this section. As a generalization to all these direct and inverse discrete sinusoidal transforms, we start from their transform formula for the k-th component, which have the following form

$$X(k,t) = \sum_{l=0}^{N-1} x(t+l)P_l(k), \qquad (5)$$

where $x(\cdot)$ is the input data sequence and $X(k,t)$ is the k-th transformed component of $\{x(t), \cdots, x(t+N-1)\}$. The weighting coefficients $P_l(k)$, $l = 0, \cdots, N-1$, are different from one transform to another. For discrete sinusoidal transforms, P_l's originate from orthogonal polynomials. From the *Fundamental Recurrence Formula*[5] for an orthogonal polynomial system, we find that $P_l = -cP_{l-1} - \lambda P_{l-2}$, where c and λ are constants for all the sinusoidal transforms that we are interested in. Using the recurrence formula, $P_l = -cP_{l-1} - \lambda P_{l-2}$, the transfer function between $X(k,\cdot)$ and $x(\cdot)$ is given by be

$$H(z) = \frac{\lambda P_{N-1} - P_N z^{-1} - \lambda P_{-1} z^{-N} + P_0 z^{-N-1}}{\lambda + cz^{-1} + z^{-2}}. \qquad (6)$$

The denominator is always a second order polynomial with the same coefficients of the recurrence relation of P_l's. When $P_0 = \pm P_N$ and $P_{-1} = \pm P_{N-1}$, which are often satisfied for the discrete sinusoidal transforms, the transfer function is further simplified by extracting out the factor $(1 + z^{-N})$ or $(1 - z^{-N})$. This extracted factor can be interpreted as an updating process and implemented by a register array. The resulting transfer function depends on λ, c, P_0, and P_{-1}, which dependent on k and the transformation. Note that the poles are always cancelled by zeros and the transfer function is essentially an FIR. In the case when two transforms can be generated dually [1], they share the same autoregressive model in their IIR filter structure. One advantage of the IIR implementation over dual generation is that it generates only one transform. When one of the transform in the dual generation is not needed, the IIR implementation is still favorable. It also can be shown that our design uses the least amount of memory asymptotically. The speed of our VLSI design cannot be improved asymptotically since it processes the input in real time. We can show that our design is asymptotically optimal in both speed and area [6].

4. CONCLUSIONS

In this paper, we present a new optimal unified architecture to compute the discrete sinusoidal transforms. The DCT, DST, DHT, DFT, LOT, and CLT are all unified under the same time-recursive, IIR implementation. To generate 1-D DCT, this architecture requires only $2N-2$ multipliers, that is one-third of the previous best known lattice structure in [1]. In Section 3, we provide a theoretical justification to show that the proposed unified filter architectures have a second order autoregressive structure. Furthermore, the throughput of this scheme is one input sample per clock cycle. The resulting architecture is regular, locally-connected and asymptotically optimal in terms of area-time complexity. This makes it very suitable for high speed real-time applications.

References

[1] K. J. R. Liu, and C. T. Chiu, "Unified Parallel Lattice Structures for Time-Recursive Discrete Cosine/Sine/Hartley Transforms," to appear IEEE Trans. on Signal processing, March 1993.

[2] B. G. Lee, "A new algorithm to compute the discrete cosine transform," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-32, pp 1243-1245, Dec. 1984.

[3] H. S. Hou, "A fast recursive algorithm for computing the discrete cosine transform," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-35, pp 1455-1461, Oct. 1987.

[4] R. Young, and N. Kingsbury, "Motion Estimation using Lapped Transforms," IEEE ICASSP Proc., pp. III 261-264, March, 1992.

[5] T. S. Chihara, *An Introduction to Orthogonal Polynomials*. Gordon and Breach, 1978.

[6] K. J. R. Liu, C. T. Chiu, R. K. Kolagotla, and J. F. JáJá, "Optimal Unified Architectures for Real-Time Computation of Time-Recursive Discrete Sinusoidal Transforms," submitted for publication.