High Precision DCT CORDIC Architectures for Maximum PSNR

Imen Ben Saad  
Université de Tunis El Manar  
Faculté des Sciences de Tunis, LAPER, UR-17-ES11, Campus Universitaire 2092, El Manar.

Sonia Mami  
Université de Tunis El Manar  
Faculté des Sciences de Tunis, LAPER, UR-17-ES11, Campus Universitaire 2092, El Manar.

Yassine Hachaïchi  
Université de Carthage, TUNISIA.  
ENICarthage  
Université de Carthage, TUNISIA.  
LAMSIN-ENIT

Younes Lahbib  
Université de Carthage, TUNISIA.  
ENICarthage  
Université de Monastir, TUNISIA.  
Electronics and Micro-electronics Laboratory

Abdelkader Mami  
Université de Tunis El Manar  
Faculté des Sciences de Tunis, LAPER, UR-17-ES11, Campus Universitaire 2092, El Manar.

Abstract—This paper proposes two optimal Cordic Loeffler based DCT (Discrete Cosine Transform algorithm) architectures: a fast and low Power DCT architecture and a high PSNR DCT architecture. The rotation parameters of CORDIC angles required for these architectures have been calculated using a MATLAB script. This script allows the variation of the angle's precision from $10^{-2}$ to $10^{-8}$. The experimental results show that the fast and low Power DCT architecture corresponds to the precision $10^{-3}$. Its complexity is even lower than the BinDCT which is a reference in terms of low complexity and its power has been enhanced in comparison with the conventional Cordic Loeffler DCT by 12 mW. The experimental results also show that the high PSNR DCT architecture corresponds to the precision $10^{-3}$ for which the PSNR has been improved by 6.55 dB in comparison with the conventional Cordic Loeffler DCT. Then, the hardware implementation and the generated RTL of some required Cordinacs are presented.

Keywords—Cordic Loeffler DCT; high quality architecture; low power architecture; Image Processing; DCT

I. INTRODUCTION

The Discrete Cosine Transform DCT was developed by Ahmed et al. in 1974 [1]. It is a robust approximation of the optimal Karhunen-Loève Transform (KLT) [2]. It has become one of the most widely used techniques of transforms in digital signal processing.

Many works deal with the optimization of the DCT architectures. Two principal axes are explored. The first one consists on the enhancement of the quality of the DCT in terms of precision measured through the Peak Signal to Noise Ratio (PSNR) [(3), (4)]. The reference in this case is the Loeffler based DCT which is the most precise architecture since it doesn’t contain approximations.

The second axe consist on improve the DCT in terms of power consumption [(5), (6), (7)]. In fact, it is well-known that DCT is one of the computationally intensive transforms since it requires many multiplications and additions. Many researches had been done on low-power DCT designs [8], [9]. As the multiplications are energy expensive operations, several algorithms are based on additions and shifts instead of multiplications.

In 2004, Jeong et al. [9] suggested improving a Cordic (Coordinate Rotation Digital Computer) based implementation of the DCT. CORDIC is an algorithm which can be used to evaluate various functions in signal processing [10], [11], [12]. In [9], authors proposed a low-complexity CORDIC based DCT algorithm based on the Flow Graph Algorithm (FGA) which is the commonly used way to represent the fast DCT. It requires only 38 add and 16 shift operations and consumes about 26.1% less power compared to [13], with a minor image quality degradation of 0.04 dB.

In the same direction, Sun et al. [14], [15] proposed a new flow graph for Cordic based Loeffler DCT implementation. A new table of parameters is obtained with new choice of the elementary rotations. Their experimental result shows that the Cordic-based Loeffler DCT consumes 16% of energy compared to [16] with a minor image quality degradation of 0.03 dB.

After this analysis of state of the art, we remark that previous works have almost neglected the quality of the results provided by the DCT algorithm in order to decrease the energy consumption. In the aforementioned works, the reached precision degree is at most $10^{-4}$. We propose to remain in the same interval ($10^{-1}$ to $10^{-4}$) and provide 2 optimal architectures. The first one is a fast and low power DCT architecture and the second one is a high PSNR DCT architecture. The parameters of the two architectures are obtained from a Matlab script which calculates the rotation parameters of the considered angles.

Contribution in this paper are:

• A matlab script which calculates the CORDIC param-
The results of the rotation iterations need to be scaled by a compensation (scale) factor $K$. This can be done by using the following iterative method.

$$K = \prod_{i} K_i = \prod_{i} \frac{1}{\sqrt{1 + 2^{-2i}}}$$  \hspace{1cm} (3)

The scale factor $K$ which can be interpreted as a constant gain (hence not data dependent) can be tolerated in many digital signal processing applications. Hence, it should be carefully investigated whether it is necessary to compensate for the scaling at all. If scale factor correction cannot be avoided, two possibilities are known. The first approach consists on performing a constant factor multiplication with $1/K_i$. The second method is based on extending the Cordic iteration in a way that the resulting inverse of the scale factor takes a value. In other words, writing the scaling factor as a sum of $2^{-i}$ where $i$ must be determined so that the error is minimized, is needed. In the rotation mode, the angle accumulator is initialized with the desired rotation angle. The rotation decision at each iteration is made to diminish the magnitude of the residual angle in the accumulator one. The decision at each iteration is therefore based on the sign of the residual angle after each step [10].

### B. Cordic-Based DCT Architecture

The One-dimensional DCT for 8x8 sub-images is defined as

$$X(t) = \frac{1}{2} C(t) \sum_{i=0}^{7} x(i) \cos \left( \frac{(2i + 1)t\pi}{16} \right)$$

$$C(t) = \begin{cases} \frac{\sqrt{2}}{2} & \text{if } t = 0 \\ 1 & \text{otherwise} \end{cases}$$  \hspace{1cm} (4)

Where $x(i)$ is the input data and $X(t)$ is 1-D DCT transformed output data.

The two-dimensional DCT is a separable transform. It can be executed by one-dimensional DCT in a serial manner as shown in the Fig. 2.

![Fig. 2. 8 × 8, 2-D DCT processor with separable 1-D DCT](IJACSA International Journal of Advanced Computer Science and Applications, Vol. 8, No. 5, 2017)

The 1-D DCT transform is represented by the Equation 5.

$$X(0) = x(0) + x(1) + x(2) + x(3) + x(4) + x(5) + x(6) + x(7)$$  \hspace{1cm} (5)

$$X(1) = \left( x(0) - x(7) \right) \cos(\pi/16) + \left( x(1) - x(6) \right) \cos(3\pi/16)$$

$$+ \left( x(3) - x(4) \right) \sin(\pi/16) + \left( x(2) - x(5) \right) \sin(3\pi/16)$$  \hspace{1cm} (6)

$$X(2) = \left( x(1) + x(6) - x(2) - x(5) \right) \cos(3\pi/8)$$

$$+ \left( x(0) + x(7) - x(3) - x(4) \right) \sin(3\pi/16)$$  \hspace{1cm} (7)
angle CORDICs and the post-scaling factors of 8-point DCT.

Fig. 3. Hardware architecture of CORDIC-based 1-D DCT

By three major components, the butterfly operator, the fixed-angle CORDICs and the post-scaling factors of 8-point DCT. Therefore, the general signal flow graph of CORDIC-based DCT is presented by Fig. 4.

\[
X(3) = (x(0) - x(7)) \cos(3\pi/16) - (x(3) - x(4)) \sin(3\pi/16) - (x(2) - x(5)) \cos(\pi/16) - (x(1) - x(6)) \sin(\pi/16)
\]

\[
X(4) = (x(0)+x(3)+x(4)+x(7)) - (x(1)+x(2)+x(5)+x(6))
\]

\[
X(5) = (x(3) - x(4)) \cos(3\pi/16) + (x(0) - x(7)) \sin(3\pi/16) - (x(1) - x(6)) \cos(\pi/16) + (x(2) - x(5)) \sin(\pi/16)
\]

\[
X(6) = (x(0) + x(7) - x(3) - x(4)) \cos(3\pi/8) - (x(1) + x(6) - x(2) - x(5)) \sin(3\pi/8)
\]

\[
X(7) = (x(0) - x(7)) \sin(\pi/16) - (x(1) - x(6)) \sin(3\pi/16) - (x(2) - x(5)) \cos(3\pi/16) - (x(3) - x(4)) \cos(\pi/16)
\]

The unfolded and reorganized equations allow to detail the origin of the FGA based DCT shown in Fig. 3. These equations are also used to represent the DCT as a matrix which will be used in the 2D-DCT processing (Fig. 2).

The Cordic array performs the fixed-angle rotation in the DCT algorithm. Therefore, the general signal flow graph of Cordic-based DCT is presented by Fig. 4.

Fig. 4. The general signal flow graph CORDIC-based DCT

Algorithm is enhancing the degree of precision by improving the selected parameters in order to find the exact values of the rotations.

\[P_{geois} \text{ Algorithm}]

The proposed MATLAB script takes as input the rotation angle. We vary the precision degree from $10^{-1}$ to $10^{-4}$ to remain in the same interval exploited by the conventional architectures.

Input Theta (angle) and Epsilon (tolerance);

The MatLab script

1: Cpt=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
2: N=1;
3: Sig=0;
4: while (abs(Theta) > Epsilon) && (N<15))
5: if (Theta < 0) Sig=Sig+1; end;
6: x=|tan(Theta)|;
7: k=round(log2(1/x));
8: Theta=|Theta|-atan(1/2^k);
9: Cpt(N)=(-1)^Sig * k;
10: N=N+1;
11: end;
12: a=0;
13: for k=1:15
14: if (Cpt(k) = 0)
15: a=a+|Cpt(k)|*atan(2^(-|Cpt(k)|))/|Cpt(k)|;
16: end
17: end

This approach provides the Cordic parameters (iterations and direction) corresponding to the angle and the selected precision. The iterations, in other words, the micro-rotations are identified with their orientation, clockwise or anticlockwise.

This method is applicable to the angles comprised within the range of 0 and $\pi/4$. The angles higher than $\pi/4$ can be decomposed into angles in this interval. For example, $3\pi/8 = \pi/4 + \pi/8$. So, to determine the CORDIC parameters of this angle, we begin by the CORDIC parameters of $\pi/4$ followed by the CORDIC parameters of $\pi/8$.

B. Cordic parameters corresponding to the angle $3\pi/16$

For a precision degree of $10^{-1}$ and $10^{-3}$, the micro-rotations shown respectively in the Table I, II are found.
The rotation angle $\frac{3\pi}{16}$ can be written as the weighted sum of micro-rotations as seen in the Equation 13

$$\theta = \frac{3\pi}{16} = 0.589048 \approx \theta_1 + \theta_3 = 0.588002 \pm 10^{-1} \quad (13)$$

Based on the previous computed micro-rotations of the $3\pi/16$ angle, the Cordic architecture computing $3\pi/16$ angle is given in Fig. 5.

The rotation angle $\frac{3\pi}{16}$ is shown in the Eq. 14

$$\theta = \frac{3\pi}{16} = 0.589048 \approx \theta_1 + \theta_3 + \theta_{10} = 0.588979 \pm 10^{-3} \quad (14)$$

The Cordic architecture computing $3\pi/16$ angle is given in Fig. 6.

C. Cordic parameters corresponding to the angle $\pi/16$

For a precision degree of $10^{-1}$ and $10^{-3}$, the micro-rotations shown respectively in the Table III and IV are found.

The rotation angle $\frac{\pi}{16}$ can be written as the weighted sum of micro-rotations as seen in the Equation 15

$$\theta = \frac{\pi}{16} = 0.196349 \approx \theta_2 = 0.244978 \pm 10^{-1} \quad (15)$$

The Cordic architecture computing $\pi/16$ angle is given in Fig. 9.
TABLE III. DETERMINING THE CORDIC PARAMETERS FOR $\pi/16$ CORRESPONDING TO A PRECISION DEGREE OF $10^{-1}$

| Iteration | $x = |tan0|$ | $i = round(-\log_2(x))$ | $\theta = |\theta| - tan^{-1}(2^{-i})$ | $\sigma$ | Stop Condition |
|-----------|---------|----------------|----------------|-------|----------------|
| 0         | 0.1989  | 2              | $0.0138$       | +     | $0.048 < 10^{-1}$ |

TABLE IV. DETERMINING THE CORDIC PARAMETERS FOR $\pi/16$ CORRESPONDING TO A PRECISION DEGREE OF $10^{-3}$

| Iteration | $x = |tan0|$ | $i = round(-\log_2(x))$ | $\theta = |\theta| - tan^{-1}(2^{-i})$ | $\sigma$ | Stop Condition |
|-----------|---------|----------------|----------------|-------|----------------|
| 0         | 0.1989  | 2              | $0.0138$       | +     | $0.048 > 10^{-2}$ |
| 1         | 0.0487  | 4              | $0.0138$       | $-$   | $0.0138 > 10^{-1}$ |
| 2         | 0.0138  | 6              | $0.0138$       | +     | $-0.0018 > 10^{-2}$ |
| 3         | 0.0018  | 9              | $0.0138$       | +     | $-1.19008e-04 < 10^{-3}$ |

IV. EXPERIMENTAL RESULTS

In order to demonstrate the high-quality feature of the proposed DCT architectures, it has been evaluated considering a JPEG2000 compression chain [18] using a well-known test image. Table VII shows the comparison of the PSNR of the proposed DCT architectures for precision degrees ranged from $10^{-1}$ to $10^{-4}$, with the other conventional DCT architectures. Checked results consider high-to-low quality compression (i.e. quantization factors from 95 to 70) using Lena image. Fig. 13 gives the experimental results based on the Lena image.

It can be easily noticed from the Table VII that Arch.Deg3 has better quality about 6.55 dB for Q=95 than the Cordic-based Loeffler. As seen in the Table VII (especially the last row which corresponds to the average PSNR), Arch.Deg3 is the closest to the Loeffler DCT which is considered as the reference and the target in terms of precision and image quality. It is also noticed that it is useless to go higher than $10^{-3}$ since the values remain stable. This is why Arch.Deg3 is considered as the best architecture in terms of image quality.

The considered architectures have been implemented on Virtex5 xc5vlx30-3ff676. The power consumption is measured with Xpower Analyzer with 100 MHz clock frequency and 1V supply power. The delay of each architecture is determined with the ISE Simulator (ISIM). The power consumption, the latency and the complexity of the different DCT architectures (the conventional and the proposed ones) with precision degrees ranged from $10^{-1}$ to $10^{-4}$ are shown in the Table VIII.

As it could be noticed, the most interesting architecture in terms of power consumption and execution delay is Arch.Deg1 which corresponds to a precision degree of $10^{-1}$. The complexity of this architecture is even lower than the BinDCT which is a reference in terms of low complexity. The power consumption of Arch.Deg1 is almost the lowest. The fact is that the power of the BinDCT is lower but this loss of power is minor when the significant enhancement made by Arch.Deg1 in terms of image quality in comparison with the BinDCT is considered.

The waveform corresponding to Arch.Deg1 and Arch.Deg3 are shown respectively in Fig. 14 and 15. As it is notable from Table VIII, Fig. 14 and 15, the execution time of a single column of an $8 \times 8$ image block is 95 ns for Arch.Deg1 and 105 ns for Arch.Deg3. In terms of number of cycles, it could be said that for Arch.Deg1 it is equal to 10 cycles and for Arch.Deg3 11 cycles. The process of an entire $8 \times 8$ image block takes 905 ns for Arch.Deg1 and 985 ns for Arch.Deg3.
TABLE V. DETERMINING THE CORDIC PARAMETERS FOR $3\pi/8$ CORRESPONDING TO A PRECISION DEGREE OF $10^{-1}$

<table>
<thead>
<tr>
<th>Iteration</th>
<th>$x$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>Stop Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration1</td>
<td>$x=1$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0 &lt; 0.1$</td>
</tr>
<tr>
<td>Iteration2</td>
<td>$x=0.4142$</td>
<td>$1$</td>
<td>$0.0709$</td>
<td>$0.95 &lt; 10^{-1}$</td>
</tr>
<tr>
<td>Iteration3</td>
<td>$x=0.0711$</td>
<td>$4$</td>
<td>$0.0085$</td>
<td>$0.07 &gt; 10^{-1}$</td>
</tr>
<tr>
<td>Iteration4</td>
<td>$x=0.0085$</td>
<td>$7$</td>
<td>$7.1738e-04$</td>
<td>$7.1738e-04 &lt; 10^{-3}$</td>
</tr>
</tbody>
</table>

TABLE VI. DETERMINING THE CORDIC PARAMETERS FOR $3\pi/8$ CORRESPONDING TO A PRECISION DEGREE OF $10^{-3}$

<table>
<thead>
<tr>
<th>Iteration</th>
<th>$x$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>Stop Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration1</td>
<td>$x=1$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0 &lt; 0.1$</td>
</tr>
<tr>
<td>Iteration2</td>
<td>$x=0.4142$</td>
<td>$1$</td>
<td>$0.0709$</td>
<td>$0.95 &gt; 10^{-1}$</td>
</tr>
<tr>
<td>Iteration3</td>
<td>$x=0.0711$</td>
<td>$4$</td>
<td>$0.0085$</td>
<td>$0.07 &gt; 10^{-1}$</td>
</tr>
<tr>
<td>Iteration4</td>
<td>$x=0.0085$</td>
<td>$7$</td>
<td>$7.1738e-04$</td>
<td>$7.1738e-04 &lt; 10^{-3}$</td>
</tr>
</tbody>
</table>

In terms of number of cycles, it could be said that Arch.Deg1 takes 80 cycles and Arch.Deg3 88 cycles. This is perfectly normal since Arch.Deg3 requires more shift/add operation layers than Arch.Deg1. So the process takes more time.

In comparison with the Loeffler DCT, it could be said that Arch.Deg1 is somewhat slower since the multiplication operation is replaced by several layers of shift/add operators which leads to a little higher delay.

If one compares the conventional Cordic Loeffler based architecture, Arch.Deg1 and Arch.Deg2, he finds that the delay is the same even though the shift/add operation layers are not exactly similar. This is perfectly normal since the delay depends essentially on the longest path and in these three cases, the longest path passes through the $3\pi/16$ Cordic.
TABLE VII. PSNR FROM HIGH-TO-LOW COMPRESSION QUALITY IN JPEG2000 FOR LENA128 IMAGE

<table>
<thead>
<tr>
<th>Quality Factor</th>
<th>Loef. DCT</th>
<th>CLoef. DCT</th>
<th>BinDCT</th>
<th>Arch.Deg1</th>
<th>Arch.Deg2</th>
<th>Arch.Deg3</th>
<th>Arch.Deg4</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>44.25</td>
<td>36.98</td>
<td>26.94</td>
<td>41.33</td>
<td>42.04</td>
<td>43.53</td>
<td>43.53</td>
</tr>
<tr>
<td>85</td>
<td>35.46</td>
<td>34.30</td>
<td>26.65</td>
<td>35.06</td>
<td>35.18</td>
<td>35.35</td>
<td>35.35</td>
</tr>
<tr>
<td>80</td>
<td>34.36</td>
<td>33.71</td>
<td>26.57</td>
<td>34.03</td>
<td>34.12</td>
<td>34.28</td>
<td>34.28</td>
</tr>
<tr>
<td>75</td>
<td>33.61</td>
<td>33.18</td>
<td>28.48</td>
<td>33.39</td>
<td>33.46</td>
<td>33.56</td>
<td>33.56</td>
</tr>
<tr>
<td>70</td>
<td>33.42</td>
<td>34.88</td>
<td>28.71</td>
<td>34.66</td>
<td>34.71</td>
<td>37.19</td>
<td>37.19</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) 34.28 dB for Q=75 (Arch.Deg3)
(b) 36.99 dB for Q=85 (Arch.Deg3)
(c) 43.53 dB for Q=95 (Arch.Deg3)
(d) 36.98 dB for Q=95 (Conventional CLDCT)

Fig. 13. Lena images obtained using the proposed CORDIC-based Loeffler DCT for P = 10^{-3}

TABLE VIII. COMPLEXITY AND POWER CONSUMPTION FOR DIFFERENT DCT ARCHITECTURES

<table>
<thead>
<tr>
<th>#point DCT</th>
<th>Multipliers</th>
<th>Add/Sub Shift</th>
<th>Power(W)</th>
<th>Delay(ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loeffler DCT [16]</td>
<td>11</td>
<td>29</td>
<td>0</td>
<td>0.744</td>
</tr>
<tr>
<td>CORDIC-based Loeffler DCT [14], [15]</td>
<td>0</td>
<td>38</td>
<td>16</td>
<td>0.642</td>
</tr>
<tr>
<td>Bin DCT [17]</td>
<td>0</td>
<td>36</td>
<td>17</td>
<td>0.600</td>
</tr>
<tr>
<td>Arch.Deg1 (10^{-3})</td>
<td>0</td>
<td>34</td>
<td>12</td>
<td>0.630</td>
</tr>
<tr>
<td>Arch.Deg2 (10^{-3})</td>
<td>0</td>
<td>40</td>
<td>18</td>
<td>0.640</td>
</tr>
<tr>
<td>Arch.Deg3 (10^{-3})</td>
<td>0</td>
<td>46</td>
<td>24</td>
<td>0.656</td>
</tr>
<tr>
<td>Arch.Deg4 (10^{-3})</td>
<td>0</td>
<td>52</td>
<td>30</td>
<td>0.659</td>
</tr>
</tbody>
</table>

Fig. 14. The Waveform corresponding to Arch.Deg1
V. Conclusion

In this paper, we present two optimal Cordic Loeffler based DCT architectures: a high PSNR architecture (Arch.Deg3) and a fast and low power architecture (Arch.Deg1). The Cordic parameters required for these architectures have been calculated using a MATLAB script. The obtained results concerning the first architecture (Arch.Deg3) show a significant improvement in the PSNR (6.55 dB for Q=95 in comparison with the Cordic Loeffler based DCT and 16.6 dB for Q=95 in comparison with the BinDCT) without a substantial loss of Power. Concerning the second architecture, we obtain an enhancement in terms of power consumption (12 mW in comparison with the conventional Cordic Loeffler based DCT and 114 mW in comparison with the Loeffler based DCT) with a significant improvement in terms of PSNR (4.35 dB for Q=95 in comparison with the Cordic Loeffler based DCT and 14.4 dB for Q=95 in comparison with the BinDCT). The optimal Cordic Loeffler DCT architectures which we found could be used in biometrical systems and endoscopy applications.

REFERENCES

[14] C.-C. Sun, S.-J. Ruan, B. Heyne, J. Goetze, 'Low-power and high quality


