Low Power 2D DCT Chip Design for Wireless Multimedia Terminals

Liang-Gee Chen, Juing-Ying Jiu, Hao-Chieh Chang, Yung-Pin Lee, and Chung-Wei Ku

DSP/IC Design Lab, Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan, R.O.C.
Tel: 886-2-363-5251 ext 443
Fax: 886-2-363-8247

e-mail: {lgchen, howard}@video.ee.ntu.edu.tw

Abstract — In this paper, a low power 2-D DCT architecture based on direct 2-D approach is proposed. The direct 2-D consideration reduces computational complexity. According to this algorithm, a parallel distributed arithmetic (DA) architecture at reduced supply voltage is derived. In the real circuit implementation of the chip, an adder of low power consumption is designed, as well as a power-saving ROM and a low voltage two-port SRAM with sequential access. The resultant 2-D DCT chip is realized by 0.6μm single-poly double-metal technology. Critical path simulation indicates a maximum input rate of 133MHz, and it consumes 138mW at 100MHz.

I. Introduction

The Discrete Cosine Transform (DCT), among various transforms, is the most popular and effective one in image and video compression, such as JPEG, MPEG, H.261 and H.263. Since these standards recently apply to battery-operated systems like portable computers (Notebook), personal digital assistants (PDA) and wireless communication equipments, it becomes imperative to develop low power DCT chip as one component of these energy-crucial desktops.

Since DCT has been standardized in recent years, many researchers and companies have took lots of resources to implement it. The conventional row-column approach has the advantage of regularity for VLSI implementation, which causes most 2-D DCT chips to be designed in this way. However, the computational complexity of the row-column approach is more than that of the direct method. And low computational amount is considered mainly in low power algorithm level. Although the direct method incurs the irregularity in realizing 2-D DCT chips, the feature of low computational complexity is still attractive for low power DCT chip design. This fact motivates our research for fewer computations and regular 2-D DCT architecture for real chip implementation with the direct method.

As to low power DCT design, T. Kuroda et al. proposed a 0.9V, 150MHz, 10mW, 2-D DCT with variable threshold-voltage scheme implemented by 0.3μm CMOS triple-well technology. However, the chip achieved low power by only taking the circuit and device level into account, not including algorithm level consideration. Therefore, we propose a 2-D DCT chip incorporating low power considerations in algorithm, architecture, and circuit design levels.

direct 2-D DCT algorithm is briefly discussed. The architecture exploiting this algorithm is described in Section III. In Section IV, The main circuit module designs, including adders and memories, are presented. The core characteristics are shown in Section V. Finally, a conclusion is given in Section VI.

II. The Direct 2-D DCT algorithm

The 2-D DCT of an N x N real signal $x_{n1,n2}$, with kernel factor $2c(n_1)c(n_2)/N$ neglected, is defined as:

$$Y_{k_1,k_2} = \sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} x_{n_1,n_2} \cdot \cos\left[\frac{2\pi(2n_1+1)k_1}{4N}\right]\cos\left[\frac{2\pi(2n_2+1)k_2}{4N}\right] \quad (1)$$
$$n_1, n_2, k_1, k_2 = 0,1,\ldots N-1$$

In the following, assume that N is to be a power of 2. Using the permutation, signal $x_{n1,n2}$ can be permuted as:

$$\begin{aligned} y_{n_1,n_2} &= x_{2n_1,2n_2} \\ &\quad n_1 = 0,\ldots,N/2-1, n_2 = 0,\ldots,N/2-1 \\ &= x_{2N-2n_1-1,2n_2} \\ &\quad n_1 = N/2,\ldots,N-1, n_2 = 0,\ldots,N/2-1 \\ &= x_{2n_1,2N-2n_2-1} \\ &\quad n_1 = 0,\ldots,N/2-1, n_2 = N/2,\ldots,N-1 \\ &= x_{2N-2n_1-1,2N-2n_2-1} \\ &\quad n_1 = N/2,\ldots,N-1, n_2 = N/2,\ldots,N-1 \end{aligned}$$

Thus, $Y_{k1,k2}$ can be rewritten as:

$$Y_{k_1,k_2} = \sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} y_{n_1,n_2} \cdot \cos\left[\frac{2\pi(4n_1+1)k_1}{4N}\right]\cos\left[\frac{2\pi(4n_2+1)k_2}{4N}\right] \quad (2)$$
$$n_1, n_2, k_1, k_2 = 0,1,\ldots N-1$$

Now consider the following expression:

$$Y_{k_1,k_2} = \sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} y_{n_1,n_2} W_{4N}^{(4n_1+1)k_1+(4n_2+1)k_2}, \qquad (3)$$

$$where \qquad W_{4N} = \exp\left(-j\frac{2\pi}{4N}\right)$$

It is not difficult to find that $Y_{kl,k2}$ can be computed from $U_{kl,k2}$ by the following set of expressions:

$$Y_{k_1,k_2} = \frac{1}{2}\left[\mathrm{Re}\left(U_{k_1,k_2}\right) - \mathrm{Im}\left(U_{N-k_1,k_2}\right)\right]$$

$$Y_{k_1,N-k_2} = \frac{1}{2}\left[-\mathrm{Im}\left(U_{k_1,k_2}\right) - \mathrm{Re}\left(U_{N-k_1,k_2}\right)\right] \qquad (4)$$

Note that (4) requires $U_{kl,k2}$ in (3) to be computed for all k_1 and only a sufficient subset of k_2 such that $\{k_2, N\text{-}k_2\}$ covers all possible values of k_2.

By the following relation [4]

$$4n_2 + 1 = (4t+1)(4n_1+1) \bmod 4N, \qquad (5)$$

Where $0 <= t, n_1, n_2 <= N\text{-}1$, the signal $y_{n1,n2}$ is mapped as $y_{n1,t}$. If n_1 is fixed, the mapping from n_2 to t is one-to-one. However, with different n_1, the mapping order is not the same.

By substituting (5) into (3), (3) can be rewritten as:

$$U_{k_1,k_2} = \sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} y_{n_1,t} W_{4N}^{(4n_1+1)[k_1+(4t+1)k_2]} \qquad (6a)$$

$$= \sum_{t=0}^{N-1}\left[\sum_{n_1=0}^{N-1} y_{n_1,t} W_{4N}^{(4n_1+1)[k_1+(4t+1)k_2]}\right] \qquad (6b)$$

$$= \sum_{t=0}^{N-1}(-j)^a\left[\sum_{n_1=0}^{N-1} y_{n_1,t} W_{4N}^{(4n_1+1)b}\right] \qquad (6c)$$

In the above deduction, we let $k_1+(4t+1)k_2 = aN+b$, where a integer and $0 <= b <= N\text{-}1$. We can find that the n_1's summation of $U_{kl,k2}$ is indeed an N-point 1-D DCT. An $N \times N$ 2-D DCT can therefore be realized by N N-point 1-D DCT's with some additions. Comparing with the row-column method which need $2N$ N-point 1-D DCT's to perform an $N \times N$ 2-D DCT, this approach with less operation complexity is more suitable for low power consideration in the algorithm level.

III. Low Power 2-D DCT Architecture

Since the direct 2-D DCT algorithm discussed above reduces the computation complexity, it is obvious that the architecture based on it shall lead to the goal of low power. The low power 2-D DCT architecture is shown in Figure 1. Since the DCT input and output is ranging from –255~255 and –2040~2040, respectively, the word-length of the input data is 9-bit and that of the output data is 12-bit. However, for convenience, the kernel factor $2c(n_1)c(n_2)/N$ is neglected in deducing the direct 2-D DCT method. Therefore, the word-length of the output data turns out to be 16-bit for covering all the output range. Besides, since the 1-D DCT computation is implemented with DA method, two-port SRAMs ordering the input and output data. Hence, 9-bit input data are fed word-serially and through the input SRAM, the data are converted into 64 bit-serial data for 2-D DCT. After these data are processed, the output SRAM changes the 64 word-parallel data to 16 bit-parallel data for next stage, usually zig-zag scan. The proposed 2D-DCT architecture with parallel 1-D DCT computation implemented in DA method is shown in Figure 2.

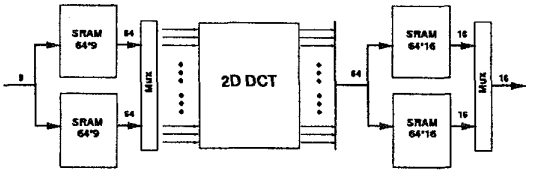


Fig. 1. Low Power 2-D DCT chip architecture

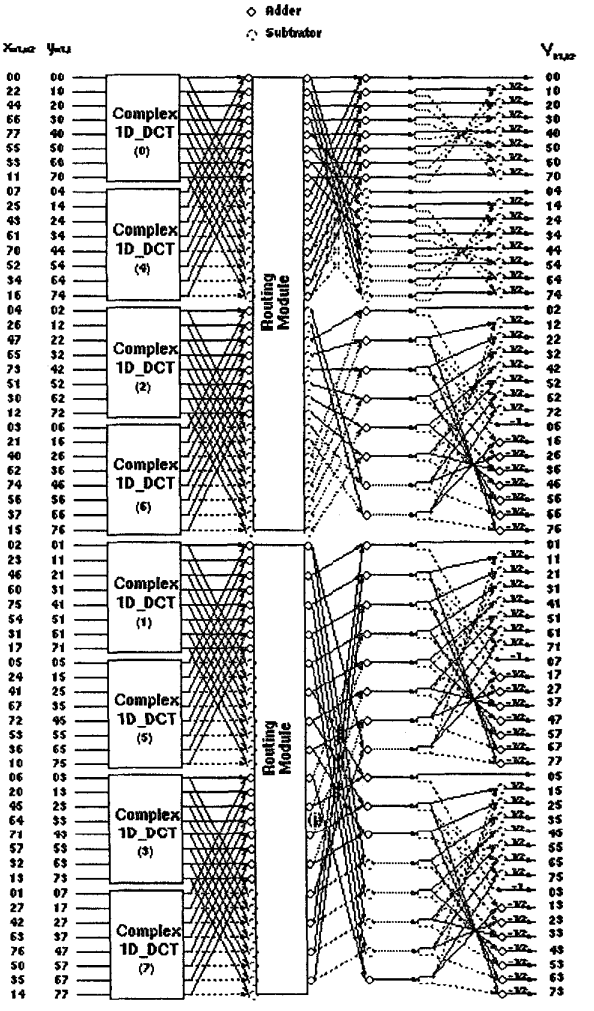


Fig. 2. The proposed parallel DA 2-D DCT architecture

IV. Chip Implementation

The proposed low power 2-D DCT chip consists of mainly adders, memories and registers. Thus, reducing the power consumption in these components will make

A. Adder Design

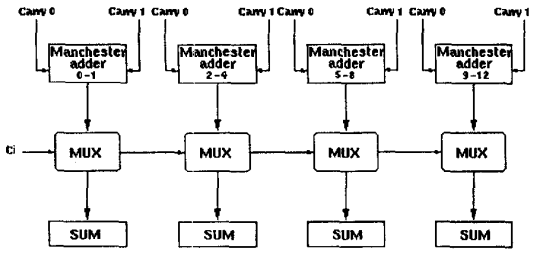
The adder is used as the accumulator in calculating the 1-D DCT result. Since the adder is also operating at low voltage, the parallelism is employed in order to compensate for the speed loss. First, the adder adopts the square-root carry-select structure shown in Figure 3. After dividing the larger adder into several stages, these stages are implemented with Manchester adder for its improvement on the carry-lookahead by using a single gate for generating carry C_i. Therefore, a large-bit adder is formed by combining the square-root carry-select adder in architecture and the Manchester adder in stage circuit. This adder has two characteristics inherited from the two adders mentioned above: carry-select for high speed and Manchester for low power.

B. Power-Saving ROM

Since the 1-D DCT in our chip is implemented by DA method, the ROM is needed to hold the content of the look-up table that is pre-computed. In order to eliminate the static power consumption due to the DC path existing in static pseudo-nMOS ROM, a better approach is to use pre-charged logic. The ROM decoder and data circuits are shown in Figure 4. An address transition detection (ATD) circuit is employed to generate the pre-charge signal *pre*, which is activated only when the input addresses change. The ROM decoder and data circuits are shown in Figure 4. During the pre-charge phase, *pre* = 0 and the bit-lines are pre-charged to V_{DD}. Meanwhile, the AND gates in decoder ensure that all pull-down paths through the NMOS are off during pre-charging. In the evaluation phase, *pre* = 1 and if the word-line is activated high, the bit-line is discharged. For the PMOS and NMOS are not turned on simultaneously during pre-charging or evaluation phase, there is no DC path from V_{DD} to *GND*, and thus, no static DC power dissipation.

C. Low-Voltage Two-Port SRAM

Since the proposed 2-D DCT is implemented with DA parallel architecture, the data reordering is needed for bit-serial word-parallel data operation. Thus, the two-port SRAM shown in Figure 5 is used for data mapping and data reordering. Note that the input port size n is different from the output port size *m*. While the two-port SRAM (*n=9, m=64*) is for the input ping-pong mode, the (*n=64, m=16*) two-port SRAM is for the ouput ping-pong mode. The sense amplifier consists of across-coupled pair of PMOS transistors and NMOS input devices. This differential pair applies the positive feedback to accelerate the sense speed.



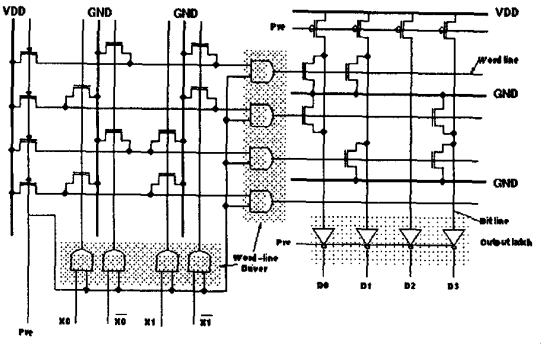


Fig. 4. ROM decoder and ROM data Circuit in the proposed power-saving ROM

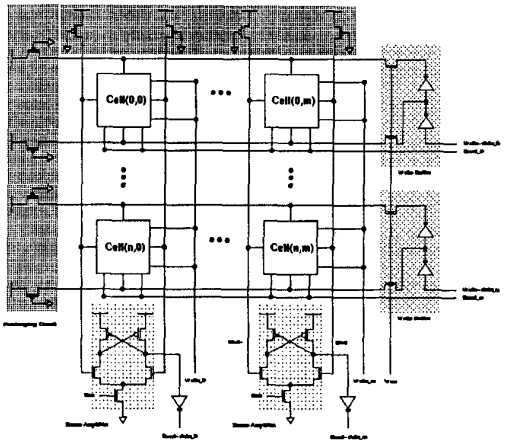


Fig. 5. The core of the two-port SRAM circuits include memory cells, write buffers, sence amplifiers and precharging circuits

V. Chip Performance and Specifications

By incorporating the module circuits discussed above, the proposed low power 2-D DCT chip with direct method is implemented. The core characteristics are summarized in Table I.

Besides, in order to understand more details about the power distribution in the designed chip, a power simulation at 100 MHz by components is shown in Table II. From this table, it is obvious that registers consume most power than others do. Then, excluding the clock buffers, the first runner up is memory module. Hence, reducing the power consumption of registers and memories will contribute more to achieve the proposed chip. That is the reason why we design low power components such as registers, memories and adders.

Since the DCT is applied to portable applications recently, the power consumption becomes a critical point in designing a 2-D DCT chip. The implementation in [1] and the product presented in [2] are not dedicated to low power design. Thus, they consume larger power. The chip reported by [3] which utilized variable threshold-voltage scheme by controlling back-bias voltage and better technology achieved a 10mW 2-D DCT core processor. The main features of these chip

Although the chip presented by [3] consumes low power, its implementation lacks the low power consideration in algorithm level. Our chip is design by taking the low power algorithm, architecture, and circuits into consideration. The ideas in both chips do not conflict. Hence, combining the low power algorithm and architecture in our chip and the variable threshold-voltage scheme in [3] will lead to a 2-D DCT chip with lower power dissipation than both two chips.

TABLE I
Chip Characteristics

Internal Word-length	16 bits
Technology	0.6μm CMOS SPDM
No. of Transistors	152017
Core Size	7.85mm 6.45mm
Die Size	8.98mm x 7.79mm
Clock Rate	100MHz
Latency	198cycles
Block Size	8 x 8
Supply Voltage	2.0 V
Power	138mW

TABLE II
Simulated Power Dissipation by Components

Module	Counts	Power(mW)	Percentage(%)
Registers	2923	35.38	25.64%
Clock buffers	1	29.35	21.27%
SRAM32x16	4	21.76	15.77%
ROM	64	17.51	12.69%
13-bit adder	64	15.48	11.22%
SRAM64x9	2	11.08	8.03%
1-bit ALU	320	5.81	4.21%
Controller	1	1.27	<0.92%

VI. Conclusion

A low-power high-performance 2-D DCT chip is implemented. The design features that contribute most to this result are as follows. First, the usage of the direct 2-D DCT algorithm reduces the 2-D DCT into 1-D DCT and some additions. Also, a fast algorithm of 1-D DCT is employed. Both of these decrease the computational complexity which means low power consumption per block operation. Besides, a parallel distributed arithmetic (DA) architecture with the direct 2-D DCT approach is proposed in order to compensate the speed loss due to the reduced internal supply voltage.

In addition to the considerations in algorithm and architecture level, low power design methdologies in logic-style and circuit level are applied to the real circuit implementation of the proposed 2-D DCT. Since adders, memories and registers are the main modules of the proposed DCT design, a power-saving in these circuits contribute to the goal significantly.

Finally, the proposed low power 2-D DCT chip with direct method is implemented. The maximum frequency simulated of the chip is 133MHz at last. It meets the requirement of the real-time HDTV signal processing for the chrominance format 4:2:0 and 4:2:2. The power simulated is 138mW at 100MHz by 0.6μm single-poly double-metal technology.

Reference

[1] D. Slawecki and W. Li,"DCT/IDCT processor design for high data rate image coding," IEEE Trans. Circuits Syst. Video Technol., vol. 2, pp. 135-146, June 1992.

[2] SGS-THOMSON Microelectronics, "2-D Discrete Cosine Transform Image Processor," product num. IMSA121.

[3] T. Kuroda et al., "A 0.9V, 150 MHz, 10-mV, 4mm^2, 2-D Discrete Cosine Transform Core Processor with Variable Threshold-Voltage (VT) Scheme. "IEEE J. Solid-State Circuit", vol. 31, no. 11, pp. 1770-1779, Nov 1996.

[4] Y. P. Lee, T. H. Chen, L. G. Chen, M. J. Chen, and C. W. Ku, "A Cost-Effective Architecture for 8 x 8 2-D DCT/IDCT Using Direct Method," IEEE Trans. Circuits Syst. Video Technol. vol 7. No 3., pp. 459-467, June 1997.

[5] J. Yuan and Svensson C.,"High-Speed CMOS Circuit Technique," IEEE J. Solid-State Circuits, vol 24, no. 1, pp. 62-70, Feb. 1989.

TABLE III
Processor Comparsion

Authors	Tech.	Core area	Trans.	Voltage	Clock rate	Power
D.Slawecki et al.[1]	2μm	72.68mm^2	67929	5V	50MHz	1W
SGS-THOMSON[2]				5V	20MHz	1.5W
T.Kuroda et al.[3]	0.3μm	4mm^2	120000	0.9V	150MHz	10mW
Our Chip	0.6μm	50.6mm^2	152017	2V	100MHz	138mW