第 34 卷 第 16 期 计 算 机 工 程 2008 年 8 月
Vol.34 No.16 Computer Engineering August 2008
·博士论文· 文章编号：1000—3428(2008)16—0007—03 文献标识码：A 中图分类号：TP303

一种低功耗 DCT 硬件结构的设计

李振伟，彭思龙，马 鸿
(中国科学院自动化研究所，北京 100080)

摘 要: 提出一种基于 CSD 编码的向量内积分布式计算结构 CDA，将其应用于二维离散余弦变换(DCT)硬件设计，利用 DCT 变换矩阵的编码特点减少设计中加法器的数量及移位累加树的带宽。该结构在 Chartered 0.13 μm 工艺库上进行设计和综合，共用了 31 528 个晶体管和 1 024 bit 存储器，具有低功耗与高性能的特点，适用于图像视频等要求低功耗、实时处理的领域。

关键词: 离散余弦变换；分布式算法；低功耗

Design of Low-power DCT Hardware Architecture

LI Zhen-wei, PENG Si-long, MA Hong
(Institute of Automation, Chinese Academy of Sciences, Beijing 100080)

【Abstract】 This paper proposes a CSD-based architecture namely CDA for computing inner product when one of the input vectors is fixed, and designs 2-D Discrete Cosine Transform(DCT) architecture with it. By optimizing the transform matrix, it reduces the number of the adders and the band of the shifter/adder. It is designed and synthesized by Chartered 0.13 μm technology, and its cost is 31 528 transistors and 1 024 bit memory. Because the architecture is with low power and high performance, it is useful for image/video compression and transmission applications.

【Key words】 Discrete Cosine Transform(DCT); Distributed Arithmetic(DA); low power

1 概述

分布式算法(Distributed Arithmetic, DA)是一种计算常系数向量内积的高效方法，它采用查找表和移位累加器来取代乘法器进行运算，具有规则的硬件结构，被大部分商用DSP芯片用于进行离散余弦变换(Discrete Cosine Transform, DCT)、DFT、数字滤波等应用。传统的DA方法预先将乘积运算中所有可能的值都存入片上ROM中，数据作为ROM地址输入，从而直接获得计算结果以此加速乘法运算。然而随着输入数据数量及中间结果精度的不断提高，ROM的尺寸也会呈指数级增长，造成硬件资源及功耗的大量消耗。人们设计了很多结构[1-2]来减小ROM的尺寸，但整体功耗仍然较高。文献[3]提出了一种基于DA算法的DCT体系结构NEDA，利用蝶形单元的运算结果取代ROM进行中间结果的获取，但其蝶形单元中加法器的数量较多，移位累加树带宽过高，造成功耗的增加。本文提出了一种新的基于CSD编码的向量内积分布式计算结构CDA，并将其应用于二维DCT变换。

2 CDA 结构的数学原理

向量内积运算为

$$\boldsymbol{Y}=\sum_{i=1}^{L}A_i\times X_i \tag{1}$$

其中，A_i 是常向量系数；X_i 是输入的数据。

其矩阵形式可表示为

$$\boldsymbol{Y}=\begin{bmatrix}A_1A_2\cdots A_L\end{bmatrix}\begin{bmatrix}X_1\\X_2\\\vdots\\X_L\end{bmatrix} \tag{2}$$

将系数 $A_1A_2\cdots A_L$ 用 CSD 码格式表示，A_i 可以表示为

$$A_i=\sum_{k=N}^{M}A_i^k2^k \tag{3}$$

其中，$A_i^k\in\{-1,0,1\}$；$P=(M-N+1)$ 表示 DA 精度。则式(2)可以表示为

$$\boldsymbol{Y}=\begin{bmatrix}2^M2^{M-1}\cdots2^N\end{bmatrix}\begin{bmatrix}A_1^M&A_2^M&\cdots&A_L^M\\A_1^{M-1}&A_2^{M-1}&\cdots&A_L^{M-1}\\\vdots&\vdots&&\vdots\\A_1^N&A_2^N&\cdots&A_L^N\end{bmatrix}\begin{bmatrix}X_1\\X_2\\\vdots\\X_L\end{bmatrix}=\begin{bmatrix}2^M2^{M-1}\cdots2^N\end{bmatrix}\begin{bmatrix}Y^M\\Y^{M-1}\\\vdots\\Y^N\end{bmatrix} \tag{4}$$

由式(4)可见，由于 $A_1A_2\cdots A_L$ 的 CSD 展开矩阵 $\boldsymbol{A}$ 具有固定参数，且只包含 3 种元素：0, 1, -1，因此只需要根据其每行的参数分别对输入数据 $X_1,X_2,\cdots,X_L$ 进行简单的加、减运算，就可以得到 $Y^M,Y^{M-1},\cdots,Y^N$ 的值，再进行相应的移位累加操作即可得到所求结果 Y。因此，利用 CDA 结构进行向量内积的计算，在硬件上只需要加、减法及移位逻辑就可以实现，如图 1 所示，其结构主要由两部分构成：模块 A 利用加、减组合逻辑产生中间结果 $Y^M,Y^{M-1},\cdots,Y^N$，模块 B 对中间结果进行移位累加，得到最终结果 Y。

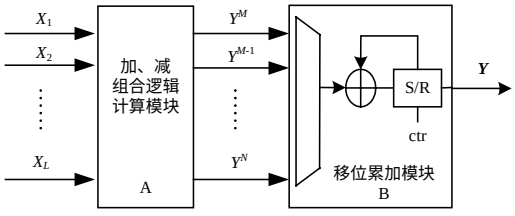

图 1 CDA 结构

基金项目：中科院自动化所青年创新基金资助项目(DG07J01)
作者简介：李振伟(1978－)，男，博士研究生，主研方向：多媒体信号处理，硬件设计；彭思龙，研究员、博士生导师；马 鸿，博士研究生
收稿日期：2007-09-30 **E-mail：**zhenwei.li@ia.ac.cn

3 基于 CDA 结构的 8×1DCT

8×8 的数据块的一维 DCT 定义为

$$Y(k)=\frac{1}{2}\sum_{i=0}^{7}C(k)x(i)\cos\frac{(2i+1)k\pi}{16} \tag{5}$$

其中，$x(i)$ 为输入数据；$Y(k)$ 是变换数据，$i,k\in[0,7]$；C 是 DCT 常系数：

$$C(k)=\begin{cases}\sqrt{1/2} & k=0\\ 1 & k\neq 0\end{cases}$$

将式(5)展开为矩阵运算。利用运算矩阵的对称性，可以分解为

$$\begin{cases}\begin{bmatrix}Y(0)\\Y(2)\\Y(4)\\Y(6)\end{bmatrix}=\begin{bmatrix}C_4 & C_4 & C_4 & C_4\\ C_2 & C_6 & -C_6 & -C_2\\ C_4 & -C_4 & -C_4 & C_4\\ C_6 & -C_2 & C_2 & -C_6\end{bmatrix}\begin{bmatrix}x(0)+x(7)\\x(1)+x(6)\\x(2)+x(5)\\x(3)+x(4)\end{bmatrix}=\boldsymbol{T}_1\cdot\begin{bmatrix}S_{1,0}\\S_{1,1}\\S_{1,2}\\S_{1,3}\end{bmatrix}\\ \begin{bmatrix}Y(1)\\Y(3)\\Y(5)\\Y(7)\end{bmatrix}=\begin{bmatrix}C_1 & C_3 & C_5 & C_7\\ C_3 & -C_7 & -C_1 & -C_5\\ C_5 & -C_1 & C_7 & C_3\\ C_7 & -C_5 & C_3 & -C_1\end{bmatrix}\begin{bmatrix}x(0)-x(7)\\x(1)-x(6)\\x(2)-x(5)\\x(3)-x(4)\end{bmatrix}=\boldsymbol{T}_2\cdot\begin{bmatrix}S_{1,4}\\S_{1,5}\\S_{1,6}\\S_{1,7}\end{bmatrix}\end{cases} \tag{6}$$

其中，$C_i=\frac{1}{2}\cos\frac{i\pi}{16}$，$i=1,2,\cdots,7$；$\boldsymbol{T}_1,\boldsymbol{T}_2$ 为变换矩阵。

显然，式(6)中存在 8 个向量内积的运算，可以分别利用 CDA 结构对其进行硬件实现。

3.1 电路设计中各模块参数的确定

在电路设计中，由于需要对 8 个DCT基进行精度为P的DA展开，而每个基的展开矩阵$\boldsymbol{A}$中第j行非零位的数量减 1 即为计算该行中间结果所需要加、减法器的数量，因此展开矩阵中的非零位越少，电路实现中需要的资源就越少，相应的计算功耗也就越少。表 1 显示了本文采用的CSD编码与NEDA[3]结构采用的二进制编码展开DCT基编码系数的比较。从表 1 可以清楚地看出，CSD编码非零位数量少于二进制编码，因此，CDA结构产生中间结果所需的资源及计算功耗都优于NEDA[3]。

表 1　CSD 码与二进制码展开对比

C_i	双精度值	2 进制编码	CSD 编码
C_1	0.490 39	0011111011000101	0100000-10-1000101
C_2	0.461 94	0011101100100000	01000-10-100100001
C_3	0.415 73	0011010100110110	010-101010100-100-1
C_4	0.353 55	0010110101000001	010-10-10101000001
C_5	0.277 79	0010001110001110	00100100-100100-10
C_6	0.191 34	0001100001111101	0010-1000100000-10
C_7	0.097 55	0000110001111100	00010-10010000-100

对于精度为P的DA展开，需要进行(P-1)次移位累加操作，P越小，需要的移位累加就越少，但P又决定了DCT精度，P越小，图像的退化越严重，因此，需要权衡计算复杂度与图像质量的关系后再决定DA展开精度P。表 2 显示了采用第 8 位~第 15 位的CSD编码以及NEDA[3]中采用 13 位二进制编码对图像进行DCT的软件仿真结果，这里采用峰值信噪比(PSNR)度量图像质量。从表 2 中可以看出，采用 11 位CSD编码的图像质量已经优于 13 位的 2 进制编码，当CSD编码精度大于 12 后，再增加编码位数对图像质量带来的好处已经很小。因此，设计中采用 12 位的DA编码精度。

表 2　不同展开精度下图像的信噪比值　　dB

	CSD8	CSD9	CSD10	CSD11	CSD12	CSD13	CSD14	CSD15	NEDA
Lena	40.239	40.241	71.988	74.123	74.602	74.501	74.616	74.596	67.518
Peppers	38.895	38.896	71.591	72.938	73.266	73.187	73.268	73.254	66.226
Bridge	44.782	44.995	64.805	74.128	78.712	77.719	79.487	79.153	70.889

3.2 具体电路结构的设计

首先，根据每个 DCT 基 CSD 展开矩阵 $\boldsymbol{A}$，进行加、减组合逻辑计算模块电路结构的设计。例如，根据式(6)，$Y(1)$ 的 CDA 展开的中间结果为

$$\begin{bmatrix}Y^0(1)\\Y^1(1)\\\vdots\\Y^{11}(1)\end{bmatrix}=\begin{bmatrix}0&0&0&0\\1&1&0&0\\0&0&1&0\\0&-1&0&1\\0&0&0&0\\0&1&1&-1\\0&0&0&0\\-1&1&0&0\\0&0&-1&1\\-1&1&0&0\\0&0&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}x(0)-x(7)\\x(1)-x(6)\\x(2)-x(5)\\x(3)-x(4)\end{bmatrix} \tag{7}$$

其中，DA 精度为 12 bit。

根据式(7)设计了如图 2(a)所示的硬件结构，利用 7 个减法器和 2 个加法器即可获得 $Y^0(1),Y^1(1),\cdots,Y^{11}(1)$ 的值。同理可得 $Y(0)$和 $Y(2)$~$Y(7)$的电路结构，如图 2(b)~图 2(h)所示。

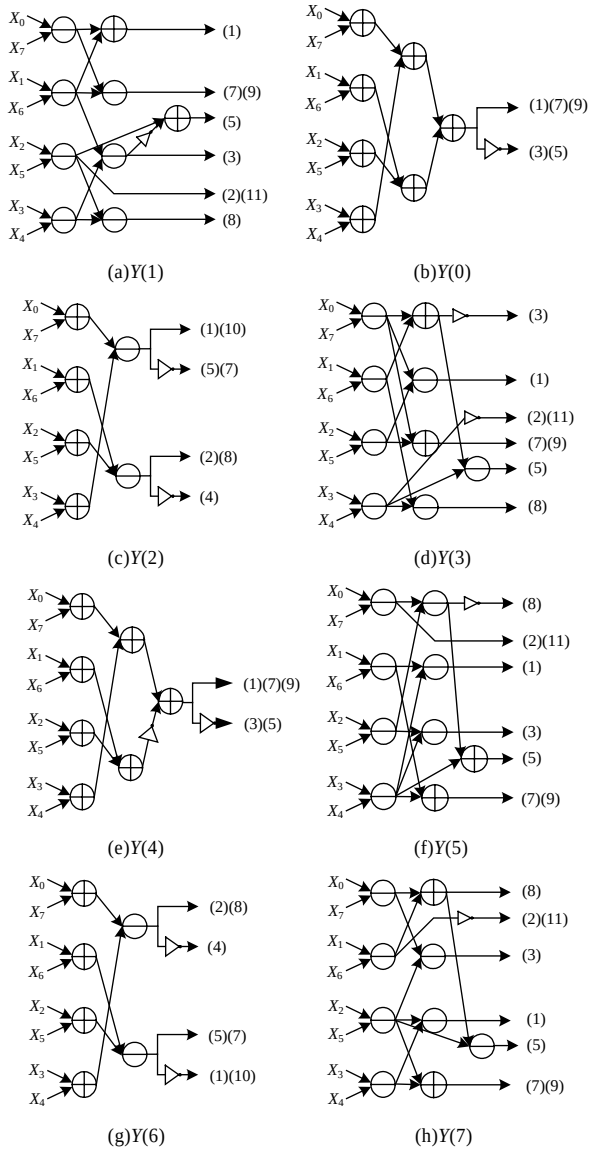

图 2　模块 A 的电路结构

如果每个 $Y(i)$都利用不同的加、减法器实现，那么模块 A

共需要 62 个加、减法运算单元。通过进一步分析可以看出，图 2 中很多计算模型是一样的，通过复用，运算单元的数量可以进一步减少。模块 A 中所有的计算模型统计如下，由此加、减法运算单元的数量减少为 26 个：

	计算模型	数量
Level 1	$S_{1,0}=X_0+X_7;S_{1,1}=X_1+X_6;S_{1,2}=X_2+X_5;$ $S_{1,3}=X_3+X_4;S_{1,4}=X_0-X_7;S_{1,5}=X_1-X_6;$ $S_{1,6}=X_2-X_5;S_{1,7}=X_3-X_4$	8
Level 2	$S_{2,0}=S_{1,0}+S_{1,3};S_{2,1}=S_{1,1}+S_{1,2};$ $S_{2,2}=S_{1,0}-S_{1,3};S_{2,3}=S_{1,0}-S_{1,3};$ $S_{2,4}=S_{1,4}+S_{1,5};S_{2,5}=S_{1,7}-S_{1,3};$ $S_{2,6}=S_{1,5}-S_{1,4};S_{2,7}=S_{1,6}-S_{1,7};$ $S_{2,8}=S_{1,4}-S_{1,6};S_{2,9}=S_{1,4}+S_{1,6};$ $S_{2,10}=S_{1,5}+S_{1,7};S_{2,11}=S_{1,6}+S_{1,7}$	12
Level 3	$S_{3,0}=S_{2,0}+S_{2,3};S_{3,1}=S_{2,0}-S_{2,3};$ $S_{3,2}=S_{1,5}+S_{2,7};S_{3,3}=S_{2,4}-S_{1,7};$ $S_{3,4}=S_{1,7}+S_{2,8};S_{3,5}=S_{1,6}-S_{2,4}$	6

然后，通过移位累加模块 B 对模块 A 的输出结果 $Y^0(1),Y^1(1),\cdots,Y^{11}(1)$ 进行移位累加，即可得到 $Y(1)$最终的值。由于每个 DCT 基展开矩阵 $\boldsymbol{A}$ 中都存在全为“0”的行，不需要进行结果累加，因此可以充分利用 $\boldsymbol{A}$ 的这个特点设计硬件电路，以减少移位累加树的带宽。

表 3 列出了 $Y(0)$~$Y(7)$需要进行移位累加的位及相应的操作次数。累加结果的移位存在左移 1 位和 2 位 2 种情况，因此，为模块 B 设计了带控制端的移位寄存器，使其可以根据控制信号对寄存器数据进行 1 位或 2 位左移操作。模块 A 将计算当前 $Y(i)$所需的数据送入模块 B 的多路器进行选通。

表 3　移位累加位统计

$Y(i)$	移位累加位(上标)	操作次数
$Y(0)$	(1)(3)(5)(7)(9)	5
$Y(1)$	(1)(2)(3)(5)(7)(8)(9)(11)	8
$Y(2)$	(1)(2)(4)(5)(7)(8)(10)	7
$Y(3)$	(1)(2)(3)(5)(7)(8)(9)(11)	8
$Y(4)$	(1)(3)(5)(7)(9)	5
$Y(5)$	(1)(2)(3)(5)(7)(8)(9)(11)	8
$Y(6)$	(1)(2)(4)(5)(7)(8)(10)	7
$Y(7)$	(1)(2)(3)(5)(7)(8)(9)(11)	8

按照上述结构进行 8×1 DCT电路设计，整体资源使用如表 4 所示。相比于NEDA[3]，本文的体系结构采用的加、减法器运算单元更少，同时移位累加树的带宽有了大幅的降低，因此，整体计算功耗有了进一步的减少。

表 4　8×1 DCT 资源使用对比

	加、减法器		移位累加树	
	数量	节省/(%)	带宽/bit	节省/(%)
直接实现	308	-	2 496	-
NEDA	35	88.64	1 936	22.44
本文实现	26	91.56	1 232	50.64

4　2D 8×8 DCT 硬件电路实现

利用二维 DCT 可分离性的特点，可以将 1 次二维 DCT 变换转化为 2 次一维 DCT 变换，如图 3 所示，利用 1D DCT 的硬件结构加入时序及控制信号以及缓存转置中间结果的 RAM 模块，同时采用三级流水处理技术，即可使二维 DCT 得到快速实现。

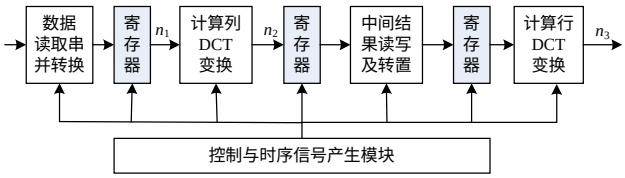

图 3　2D DCT 整体结构

上述 2D DCT变换电路的变换精度由 2 个参数决定：(1)DCT基的DA精度P；(2)数据的表示精度。在目前的国际标准中，DCT的输入范围为[-255, 255]，输出范围为[-2 048, 2 048]，因此，对于输入数据和输出数据，采用“n_1=9, n_3=12”的表示精度。由于在RAM中数据采用“$i.f$”的格式进行存储，即“i”位的整数部分加“f”位的小数部分，对于整数部分，采用“i=12”的精度已经足够；对于小数部分，f越大，变换精度就越高，但也会造成RAM尺寸的指数级增大以及数据通路的增加，因此对上述两方面进行折中，对于经过第 1 次 8×1 DCT变换后得到的中间结果存储到RAM的数据精度n_2，最终选择“f=4”即“n_2=16”的表示精度。采用上述精度对本文体系结构进行仿真，结果如表 5 所示。可以看出，它完全符合文献[4]描述的精度规范。

表 5　IEEE 1180-1190 标准及测试结果

项目	标准	本文体系结构
Pixel Peak Error(PPE)	≤ 1	1
Pixel Mean Square Error(PMSE)	≤ 0.06	0.013
Pixel Mean Error(PME)	≤ 0.015	0.011
Overall Mean Square Error(OMSE)	≤ 0.02	0.002
Overall Mean Error(OME)	≤ 0.001 5	0.000 04
All input data is 0(ALL INPUT 0)	ALL OUTPUT 0	ALL OUTPUT 0

5　实验结果与比较

采用 Verilog HDL 对第 4 节所设计的 2D DCT 变换硬件电路进行行为级描述，在 Modelsim 6.0 环境下进行功能仿真验证，并用 Synopsys 公司的 Design Compiler 和 Chartered 0.13 μm 工艺库对设计进行综合。表 6 给出了综合后的结果以及与其他几种 DCT 体系结构的比较。

表 6　本文结构与其他 DCT 结构的比较

	文献[1]结构	文献[2]结构	文献[3]结构	本文结构
实现方法	基于 ROM 的 DA	基于 ROM 的 DA	NEDA	CDA
工艺	0.25 μm	0.18 μm	0.18 μm	0.13 μm
加法器数量	32	16	35	26
存储器	4 096 bit ROM 1 024 bit RAM	2 048 bit ROM 1 024 bit RAM	1 024 bit RAM	1 024 bit RAM
晶体管数	73 000	112 000	41 300	31 528
吞吐率/(兆像素·s^{-1})	14.3	36	>83	100
功耗	Very High	Very High	Low	Low

由于采用了三级流水结构，关键路径上最长延时仅为三级进位选择加法器的延时，因此本文结构最高的时钟频率可以达到 1.5 GHz，硬件开销只有 31 528 个晶体管(7 882 门)，在满足各种实时处理应用的每秒 100 兆像素的吞吐率下，整个电路功耗仅为 236 mW，功耗与性能相对于已有的设计都有明显的改善。

6　结束语

DCT 作为图像和视频压缩的重要组成部分，被众多国际标准(如 JPEG, MPEG 和 H.26x)所采用。由于需要进行大量的乘累加运算，因此计算复杂度及功耗都很高。本文提出了一种新的基于 CSD 编码的二维 DCT 体系结构，无需 ROM 和乘法器，仅利用面积开销较低的加、减法器、移位器和多路器实现乘法计算非常密集的 DCT。利用变换矩阵的特性进行硬件设计，极大地减少了设计中加减法器的数量以及移位累加数的带宽。最终的硬件实现表明，本设计共消耗了 31 528 个晶体管和 1 024 bit 存储器，在每秒 100 兆像素的吞吐率时，功耗仅为 236 mW，具有较高的变换精度，适用于图像视频等要求低功耗、实时处理领域。

(下转第 18 页)