



一种高速近似乘法器设计

吴德祥, 班 恬

(南京理工大学 电子工程与光电技术学院, 南京 210094)

摘 要: 近似计算作为一种有效权衡精度与性能的新型计算方式, 已被广泛运用于图像处理、数据挖掘和多媒体技术等能够容忍少量计算错误的相关应用中, 然而此类应用存在大量乘法操作。为加快数据处理速度, 设计一种新型的近似乘法器, 采用近似加法实现部分累加运算, 从而减少近似乘法器的资源消耗, 同时通过流水线结构增加系统的时钟频率, 进而提高数据吞吐率。统计结果表明, 与精确乘法器相比, 该设计可节省 32.2% 的查找表资源。在图像处理应用中, 相较 AMA、UDM 等近似乘法器, 该设计的峰值信噪比较高, 图像重构的效果较好。

关键词: 近似计算; 容错; 乘法器; 流水线; 图像处理

开放科学(资源服务)标志码(OSID):



中文引用格式: 吴德祥, 班恬. 一种高速近似乘法器设计[J]. 计算机工程, 2019, 45(12): 289-293.

英文引用格式: WU Dexiang, BAN Tian. A high speed approximate multiplier design[J]. Computer Engineering, 2019, 45(12): 289-293.

A High Speed Approximate Multiplier Design

WU Dexiang, BAN Tian

(School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

[Abstract] Approximate compute, as a new calculation method that effectively balances accuracy and performance, has been widely used in applications such as image processing, data mining, and multimedia technology that can tolerate a small number of computational errors. There is the large number of multiply-accumulate operations in these applications. To speed up data processing, this paper designs a new type of approximate multiplier. The multiplier uses approximate addition to achieve partial accumulation, thus reducing the resource consumption of the approximate multiplier. Besides, a pipeline structure is adopted to increase the clock frequency of the system, thereby, increasing the data throughput. The statistical results show that compared with that of the precision multiplier, the Look-Up-Table (LUT) resource savings of the approximate multiplier reach 32.2%. Also, in image processing applications, compared with AMA, UDM approximate multipliers, the proposed multiplier has higher Peak Signal-to-Noise Ratio (PSNR) and better effect on image reconstruction.

[Key words] approximate compute; error-tolerant; multiplier; pipeline; image processing

DOI: 10.19678/j.issn.1000-3428.0052582

0 概述

由于人类感官的固有局限性, 如对图像、声音的细微变化不敏感, 因此诸如图像变换、音频处理等许多应用就可以采用近似计算^[1], 而无需得到精确的计算结果。近似计算的关键思想是通过牺牲少量的精度来换取性能的大幅提升。在许多应用中存在乘法运算, 采用近似乘法器可以提升处理速度, 减少资源

消耗。本文在研究现有近似乘法器的基础上, 设计一种新的高速近似乘法器, 并将其应用到图像处理中。

1 相关工作

乘法器对于给定 2 个 n 位的操作数进行乘法操作, 其主要步骤为:

1) 将操作数按位相与, 产生一系列 n 位部分乘积项 (Partial Product, PP)^[2]。

基金项目: 国家自然科学基金(61401205); 江苏高校“青蓝工程”。

作者简介: 吴德祥(1994—), 男, 硕士研究生, 主研方向为近似算术电路设计; 班恬(通信作者), 副教授。

收稿日期: 2018-09-06 修回日期: 2018-11-19 E-mail: 18362961858@163.com

2) 通过一些特定方式如全加、4-2 压缩^[2]进行累加,将 n 位 PP 压缩为 2 行。

3) 对最后 2 行求和,得到最终乘积。

乘法器的近似可以从上述 3 个步骤入手。文献[3]提出一种根据设计的乘法器(Under Designed Multiplier, UDM),通过 2 乘 2 近似乘法器子模块来产生 PP。

对 PP 压缩是近似研究的重点方向,主要方法包括截断、4-2 压缩^[4-5]以及近似全加器累加等。文献[4-7]提出不同的 4-2 压缩方法来压缩 PP。文献[8-9]利用或门来压缩低位的 PP,高位用精确加法器累加避免误差距离过大。

除了上述近似方法之外,文献[10]使用米切尔对数算法及其改进形式来近似乘法操作,获得的改进乘法器性能较好。

2 近似乘法器结构

2.1 近似加法结构

对绝大部分乘法器来说,加法器是基本元素。低位或门加法器(Lower-part-OR Adder, LOA)与容错加法器(Error-Tolerant Adder, ETA)可以对加法器低位进行不精确计算。文献[11]用模块化的思想来实现近似,除了减少延迟外,还可以通过改变模块的规模来控制加法器的精度。文献[12]提出近似镜像加法器(Approximate Mirror Adder, AMA)的概念,通过删除晶体管达到简化逻辑结构的目的。

本文对 AMA^[12]的真值表进行修改,结果如表 1 所示。近似全加器电路结构如图 1 所示。

表 1 近似全加器真值表

a	b	cin	s	$cout$
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	0
1	0	0	1	1
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

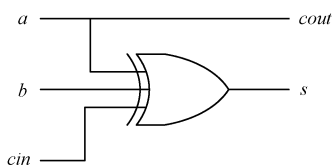


图 1 近似全加器电路结构

近似全加器的结构较精确全加器而言大为简化,在减少资源消耗方面具有显著效果。基于上述近似全加器,本文提出一种新型的近似加法结构,如图 2 所示。当给定 2 个输入操作数 $a[n-1]$ 与 $b[n-1]$ 之后,各级的进位输入也同时确定,和 $s[n-1]$ 的每位可以同时计算得出,高位结果的计算无需等待低位计算的进位输出,计算速度能够得到极大提升。由该加法结构实现的多位近似加法器具有不亚于超前进位加法器(Carry Look-ahead Adder, CLA)的高速性能,而无需计算各级的进位输出则使得该加法结构能够节省较多的电路资源。

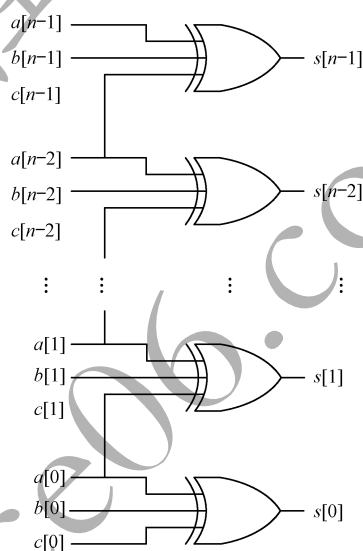


图 2 近似加法结构电路图

2.2 流水线结构

流水线被广泛应用于高速数字系统设计中^[13]。具有流水结构的设计能够在不增加过多硬件代价的前提下有效提高数字电路的数据处理速度,其基本思想是将复杂的数字逻辑电路分级实现,使每一级的电路结构简化,在较小的时钟周期内完成电路的功能^[14]。本文设计的近似乘法器(以 8 位为例)的流水线结构如图 3 所示。对一般情况而言,当输入为 2 个 n 位二进制数时,该乘法器首先通过与门生成 n 行 n 位的 PP,并将其寄存至长度为 $2n$ 的寄存器 pp0 ~ pp($n-1$) 中。定义近似长度为 l (l 值一般为 $n/4$, $n/2$ 或 $3n/4$),对权重较小的 l 行 PP 利用 $2n$ 位近似加法器累加。然后将近似部分与精确部分的和分别寄存至 $2n$ 位寄存器 APP 与 ACC 中。最后利用 $2n$ 位 CLA 将近似部分的和与精确部分的和相加,得到最终结果。当时钟沿到来时,上述 3 个操作过程同时工作,可以提高系统的时钟频率,进而提升数据吞吐率。

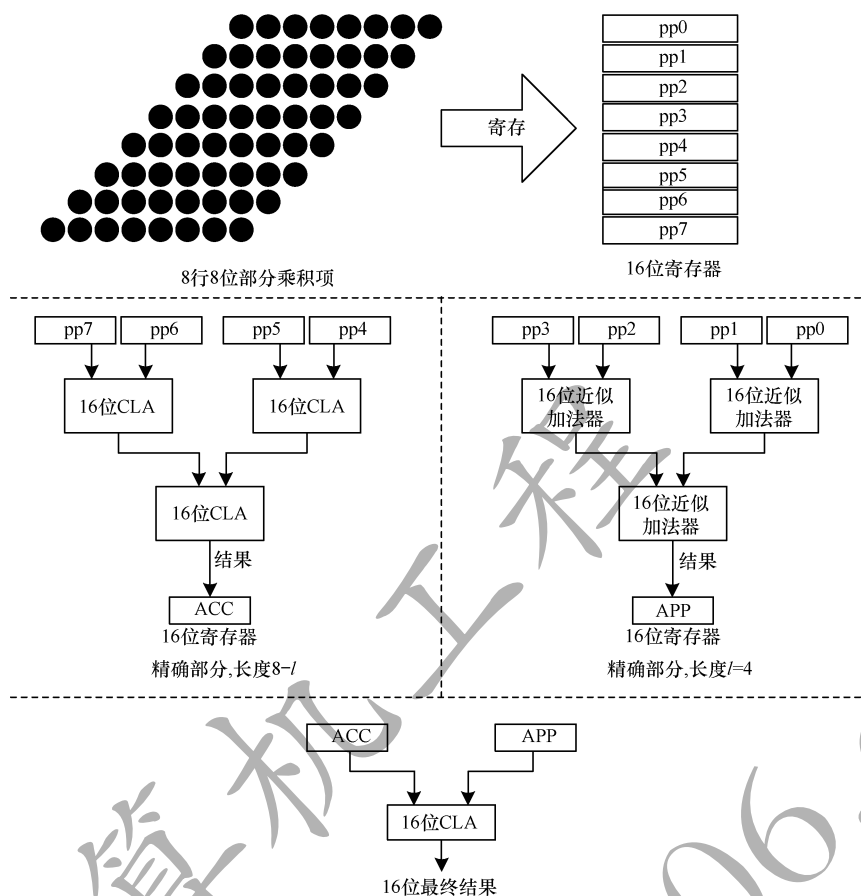


图3 近似乘法器流水线结构

3 性能评估

3.1 误差特性

误差距离 (Error Distance, ED) 表示精确结果与近似结果之差的绝对值。平均误差距离 (Mean Error Distance, MED) 表示一组 ED 的均值。归一化平均误差距离 (Normalized MED, NMED) 统一了不同位宽的近似乘法器的 MED 对比标准。错误率 (Error Rate, ER) 表示错误的输出结果占有所有输出结果的百分比。上述参数可以有效表示近似乘法器的误差特性。

对于本文所设计的近似乘法器,近似部分的长度(即 l 值)可以进行配置,以实现乘法器的精度可控。为方便对比,本文将基于 AMA^[12] 的近似乘法器命名为设计 1,将基于 UDM^[13] 的近似乘法器命名为设计 2,将基于 4-2 压缩^[2] 的近似乘法器命名为设计 3。本文设计以 $l=4$ 为例。不同近似乘法器的误差特性如表 2 所示。其中,ED 为实验的最大误差距离即 ED 的最大值。

表 2 不同乘法器的误差特性

乘法器	ED	MED	NMED	ER/%
设计 1	2 555	732.95	0.011 3	93.38
设计 2	3 432	334.13	0.005 1	61.11
设计 3	1 441	297.22	0.004 6	88.21
本文设计	3 400	504.10	0.007 8	69.46

3.2 性能参数

● 将精确乘法器、设计 1、设计 2、设计 3 以及本文设计应用于图像处理中,统计最大延时以及所占用的查找表 (Look-Up-Table, LUT) 与触发器 (Flip-Flop, FF) 资源。以最大延时、LUT 以及 FF 三者之间的乘积 (Delay-LUT-FF Product, DLFP) 作为主要参考系数。DLFP 系数越小,表示乘法器的延时与资源消耗的总性能越好。统计结果如表 3 所示。

表 3 各乘法器资源占用对比

乘法器	最大 延时/ns	LUT 数量	FF 数量	DLFP /10 ⁹	LUT 节省 百分比/%
精确乘法器	11.586	20 810	9 401	2.267	—
设计 1	11.009	14 631	9 795	1.578	29.69
设计 2	11.578	18 489	9 547	2.044	11.15
设计 3	11.131	14 258	9 323	1.480	31.48
本文设计 ($l=4$)	8.722	18 391	9 907	1.589	11.62
本文设计 ($l=6$)	7.777	14 109	9 907	1.087	32.20

从表 3 可以看出,与其他近似乘法器相比,本文设计的最大延时最小;当 $l=6$ 时,本文设计 LUT 节省百分比达 32.20%。虽然流水线的使用使整个设计增加了少量触发器资源,但就 DLFP 系数而言,本文设计在 $l=6$ 时远优于其他设计。

4 近似乘法器应用

将本文设计应用于 DCT/IDCT 图像变换^[15]中以验证其实用性。矩阵相乘是 DCT/IDCT 的主要过程^[16]。在 FPGA 中灰度矩阵与变换矩阵相乘主要由调用乘法器 IP 核来实现。本文利用基于近似乘法器所设计的近似有符号乘法器 IP 核来替换精确的 IP 核,得到近似的 DCT/IDCT 结果。

为衡量图像重构质量,本文采用峰值信噪比(Peak to Signal Noise Ratio, PSNR)作为评估参数。本文设计与其他乘法器对不同图像作 DCT/IDCT 处理所得到的结果如图 4 所示。各重构图像的峰值信噪比如表 4 所示。

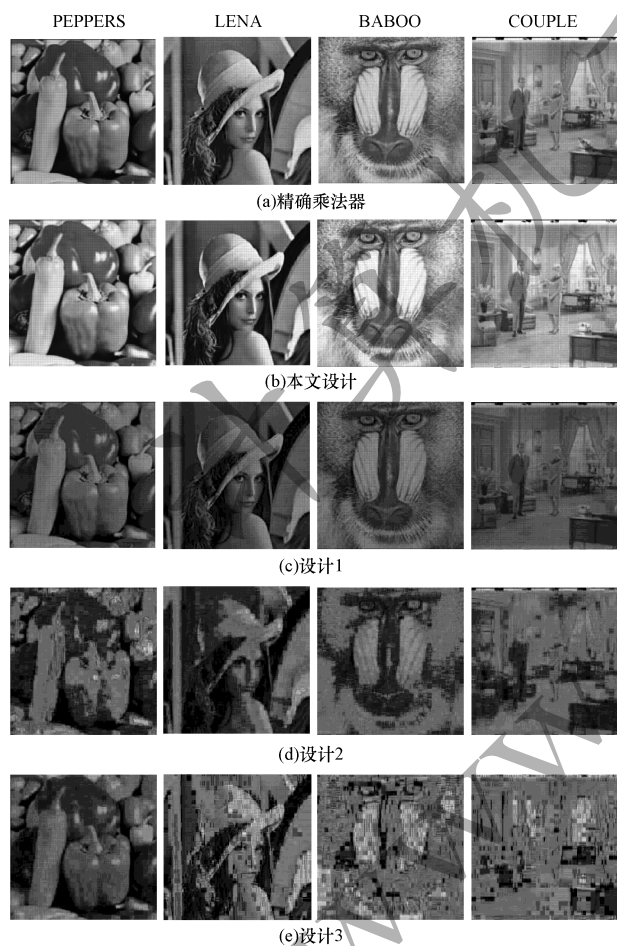


图 4 各乘法器重构图像对比

表 4 各乘法器重构后图像的 PSNR 值

图像	精确乘法器	本文设计 ($I=4$)	设计 1	设计 2	设计 3
PEPPERS	26.953 5	19.191 7	14.491 3	13.262 6	13.092 8
LENA	27.307 1	20.623 2	15.642 7	14.859 7	13.606 2
BABOO	23.786 4	18.341 0	13.870 2	13.277 7	14.320 2
COUPLE	26.396 8	19.482 7	14.445 5	15.004 7	15.413 1

从表 4 可以看出,不同图像经本文设计处理之后仍具有相当高的 PSNR 值,图像的质量损失很小。对于同一幅图,与其他近似乘法器处理之后的图像相比,本文设计处理后的图像 PSNR 值最大,图像重构的效果最好。

5 结束语

本文设计一种近似乘法器,在牺牲一定计算精度的前提下,极大提升了乘法器的计算性能,且其精度可以按照实际要求调节,能够减少资源消耗和延迟。在 DCT/IDCT 图像变换中的实验结果表明,本文乘法器重构图像的 PSNR 损失为 6 dB ~ 7 dB,质量损失相对较小。下一步将对其进行优化,在保证实用性的同时,进一步减少资源消耗及时延。此外,还将研究不同的近似设计方式,如对逻辑综合过程的近似等。

参考文献

- [1] YANG Tongxin, UKEZONO T, SATO T. A low-power high-speed accuracy-controllable approximate multiplier design [C]//Proceedings of the 23rd Asia and South Pacific Design Automation Conference. Washington D. C., USA:IEEE Press,2018:605-610.
- [2] HA Minho, LEE S. Multipliers with approximate 4-2 compressors and error recovery modules [J]. IEEE Embedded Systems Letters,2017,10(1):6-9.
- [3] KULKARNI P, GUPTA P, ERCEGOVAC M. Trading accuracy for power with an underdesigned multiplier architecture [C]//Proceedings of the 24th International Conference on VLSI Design. Washington D. C., USA: IEEE Press,2011:346-351.
- [4] SHAFIQUE M, HAFIZ R, REHMAN S, et al. Cross-layer approximate computing: from logic to architectures [C]//Proceedings of the 53rd Annual Design Automation Conference. New York, USA: ACM Press,2016:99.
- [5] BHARDWAJ K, MANE P S, HENKEL J. Power-and area-efficient approximate wallace tree multiplier for error-resilient systems [C]//Proceedings of the 15th International Symposium on Quality Electronic Design. Washington D. C., USA: IEEE Press,2014:263-269.
- [6] VENKATACHALAM S, KO S B. Design of power and area efficient approximate multipliers [J]. IEEE Transactions on Very Large Scale Integration Systems, 2017,25(5):1782-1786.
- [7] ZERVAKIS G, TSOUUMANIS K, XYDIS S, et al. Design-efficient approximate multiplication circuits through partial product perforation [J]. IEEE Transactions on Very Large Scale Integration Systems, 2016,24(10):3105-3117.

- [8] QIQIEH I, SHAFIK R, TARAWNEH G, et al. Energy-efficient approximate multiplier design using bit significance-driven logic compression [C]//Proceedings of the Conference on Design, Automation and Test in Europe. New York, USA: ACM Press, 2017:7-12.
- [9] LIU Cong, HAN Jie, LOMBARDI F. A low-power, high-performance approximate multiplier with configurable partial error recovery [C]//Proceedings of 2014 Design, Automation and Test in Europe Conference and Exhibition. Washington D. C., USA: IEEE Press, 2014:1-4.
- [10] KIM M S, DEL BARRIO A A, HERMIDA R, et al. Low-power implementation of Mitchell's approximate logarithmic multiplication for convolutional neural networks [C]//Proceedings of the 23rd Asia and South Pacific Design Automation Conference. Washington D. C., USA: IEEE Press, 2018:617-622.
- [11] KAHNG A B, KANG S. Accuracy-configurable adder for approximate arithmetic designs [C]//Proceedings of the 49th Annual Design Automation Conference. New York, USA: ACM Press, 2012:820-825.
- [12] GUPTA V, MOHAPATRA D, PARK S P, et al. IMPACT: imprecise adders for low-power approximate computing [C]//Proceedings of the 17th IEEE/ACM International Symposium on Low-power Electronics and Design. Washington D. C., USA: IEEE Press, 2011: 409-414.
- [13] ZHAO Guoliang. Research and implementation of 1024-point pipeline FFT algorithm based on FPGA [D]. Xi'an: Xidian University, 2011. (in Chinese)
赵国亮. 基于FPGA的1024点流水线结构FFT算法的研究与实现[D]. 西安: 西安电子科技大学, 2011.
- [14] FEI Yongzhou. Research and design of 12-bit high speed pipeline ADC [D]. Nanjing: Southeast University, 2017. (in Chinese)
费永舟. 12位高速流水线ADC的研究和设计[D]. 南京: 东南大学, 2017.
- [15] WU Ying. Application of DCT transform in image compression [J]. Computer and Modernization, 2013(4): 103-106. (in Chinese)
武瑛. DCT变换在图像压缩中的应用[J]. 计算机与现代化, 2013(4): 103-106.
- [16] XU Zhen. VLSI design of DCT/IDCT IP core based on image processing application [D]. Wuhan: Huazhong University of Science and Technology, 2004. (in Chinese)
薛峥. 基于图像处理应用的DCT/IDCT IP核的VLSI设计[D]. 武汉: 华中科技大学, 2004.

编辑 刘盛龄