

DOI:10.19651/j.cnki.emt.2209538

## 自适应二进制算术编码的动态可重构实现研究\*

刘尧<sup>1</sup> 蒋林<sup>1</sup> 李远成<sup>2</sup> 山蕊<sup>3</sup>(1. 西安科技大学电气与控制工程学院 西安 710600; 2. 西安科技大学计算机科学与技术学院 西安 710600;  
3. 西安邮电大学电子工程学院 西安 710121)

**摘要:** 针对 H.266/VVC 视频编码标准下的上下文自适应二进制算术编码器编码速度慢、资源开销大的问题,面向可重构结构依据算法的内在并行特性优化了编码架构,并基于动态可重构阵列处理器设计实现了 CABAC 编码器常规编码模式下的并行映射方法,阵列结构能够根据编码输入对优化后的算法进行动态重构,在避免专用硬件编码器较高的资源开销情况下利用软件重构的方法实现熵编码过程,保证编码准确性的同时提高了视频数据流编码效率,为此类运算密集型算法的硬件实现提供了更为灵活高效的参考途径。仿真结果表明,映射实现的编码过程中每个编码周期完成 5 个二进制序列的编码,平均编码效率达到 384.13 Mbin/s。基于 FPGA 的测试结果表明,软件重构方法与专用硬件实现的编码器相比,资源开销降低且编码效率提升 5.47%,与同类型可重构视频编码结构相比,编码效率提升 7.03%。

**关键词:** 可重构计算;自适应二进制算术编码;并行映射;阵列处理器;熵编码

**中图分类号:** TP919.81 **文献标识码:** A **国家标准学科分类代码:** 520.3020

## Research on dynamic reconfigurable implementation of context adaptive binary arithmetic coding

Liu Yao<sup>1</sup> Jiang Lin<sup>1</sup> Li Yuancheng<sup>2</sup> Shan Rui<sup>3</sup>(1. School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710600, China;  
2. School of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710600, China;  
3. School of Electronic Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

**Abstract:** In response to the slow coding speed and high resource overhead of the Context-based adaptive binary arithmetic coding under the H.266/VVC video coding standard, the reconfigurable oriented architecture optimized the coding architecture based on the intrinsic parallelism of the algorithm, and designs and implements a parallel mapping method for the CABAC encoder in conventional coding mode based on a dynamically reconfigurable array processor. The array structure is able to dynamically reconfigure the optimized algorithm according to the coding input, and the software reconfiguration method is used to implement the entropy coding process without the high resource overhead of dedicated hardware encoders. Simulation results show that the mapped encoding process completes five binary sequences per encoding cycle with an average encoding efficiency of 384.13 Mbin/s. FPGA based test results show that the software reconstruction approach reduces the resource overhead and improves the encoding efficiency by 5.47% compared to dedicated hardware implementations, and improves the encoding efficiency by 7.03% compared to similar reconfigurable video encoding structures.

**Keywords:** coarse-grained reconfigurable architecture; context-based adaptive binary arithmetic coding; parallel mapping; array processor; entropy coding

## 0 引言

熵编码是一种利用信源随机过程统计特性的无损编码

方式<sup>[1]</sup>,其中基于上下文的自适应二进制算术编码算法(context-based adaptive binary arithmetic coding, CABAC)是新一代通用视频编码的关键流程<sup>[2]</sup>,算法中存

收稿日期:2022-04-08

\* 基金项目:国家自然科学基金重点项目(61834005/F0402)、陕西省自然科学基金(2020JM-525)、榆林市科技计划项目(CXY-2020-026)资助

在的数据依赖影响专用硬件中流水线和并行化的实施,是限制 CABAC 硬件编码器编码速度及数据吞吐量的首要问题<sup>[3]</sup>。

编码器是最常见 CABAC 算法专用硬件实现形式<sup>[4]</sup>。此类面向特定需求的专用硬件虽然能满足算法实现的需要,但其结构固定且优化手段单一,不能根据不同编码需求动态的改变硬件实现方案,专用硬件的灵活性明显不足<sup>[5]</sup>。文献[6]设计的编码器采用并行流水线架构形式,在实现了较高数据吞吐量的同时增加了硬件资源消耗。文献[7]设计了一种基于近似插值的硬件结构,降低了逻辑资源占用率。但是,对编码过程的简化使得编码准确性有所下降。文献[8]基于重归一化对编码算法进行改进,并设计了基于流水线的编码结构,但是,编码速度有待提高。文献[9]设计采用了高度并行硬件结构和查找表优化方法,但是,该结构逻辑复杂,硬件资源开销大。文献[10-11]的硬件结构在改进概率估计过程的同时提高了编码压缩效率。但是,其资源开销大,难以针对不同编码需求进行灵活切换。由此可见,现有的熵编码结构受到专用硬件实现方式的限制,其资源开销大、带宽要求高且算法灵活性较差<sup>[12]</sup>。

论文针对专用硬件编码器的性能缺陷,基于动态可重构阵列处理器,利用软件重构的方法替代编码器等专用硬件结构,实现了上下文自适应二进制算术编码流程。在保证数据吞吐量和编码准确性的前提下,提高了 CABAC 编码速率,解决了专用硬件资源开销大,带宽要求高,不能灵活适应各种改进视频算法的问题。依靠动态可重构阵列结构的灵活性特征,本文为此类运算密集型典型算法的硬件实现提供了更为高效的参考途径,能够满足数据吞吐量大,运算复杂度高的硬件实现需求。

### 1 CABAC 编码算法并行性分析

CABAC 编码包括二值化,上下文建模以及算术编码等主要步骤,如图 1 所示。

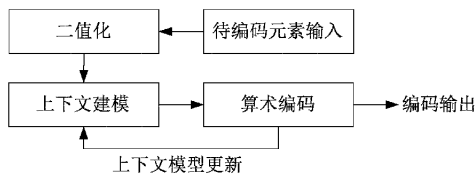


图 1 字符特征

在二值化阶段,资源开销主要集中在二进制序列 bins 存取过程,数据并行存取能够有效提高二值化效率;上下文建模对序列中每个元素出现概率进行估计,对多个 bin 并行操作,包括更新当前概率状态索引  $\sigma$  和最大、最小概率符号 MPS, LPS 等,能够有效缩短概率计算时间。

在上下文建模阶段,概率状态估计值 state 由一组有限集合表示,代表当前符号所对应的 LPS 出现概率。CABAC 概率估计更新模型如图 2 所示,通过并行化的概

率更新估计能有效节约构建编码过程所需查找表的时间。

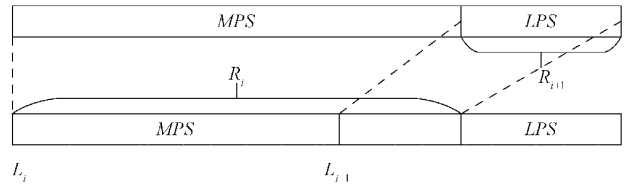


图 2 概率更新估计

在算术编码阶段,通过并行化方法将当前待编码符号 bin, LPS 值以及更新后的概率状态索引  $\sigma$  等相关输入参数提前传输到对应处理结构的数据缓存区域。描述编码区间的参数由当前区间下界 low 和当前区间长度 range 构成。

若输入的待编码符号  $bin = LPS$ , 则编码区间更新为:

$$\begin{cases} range = rLPS \\ low = low + range - rLPS \end{cases} \quad (1)$$

若  $bin = MPS$ , 则编码区间更新为:

$$\begin{cases} range = range - rLPS \\ low = low \end{cases} \quad (2)$$

编码区间更新完成后,对编码数据实时打包完成编码输出。

并行化的数据存取及编码方式能够提高编码效率并加速编码过程。同时,映射实现的残差数据存储方式有利于编码过程中查找表的构建与使用。由于输入数据块大小的不同使得专用硬件的资源占用较多,能够通过可重构阵列结构动态的适应不同块大小带来的编码算法变化。由此可见,CABAC 算法各流程具有较强的内部并行特性,基于可重构阵列的软件映射方法更便于编码结构的并行化。因此,动态可重构阵列对 CABAC 算法的主要流程均有较高的适配性。

### 2 面向可重构阵列的 CABAC 算法优化

根据动态可重构阵列结构的并行运算特性,二进制算术编码映射优化方案如图 3 所示,CABAC 编码实际映射过程分为取数、遍历、统计、编码 4 个步骤。

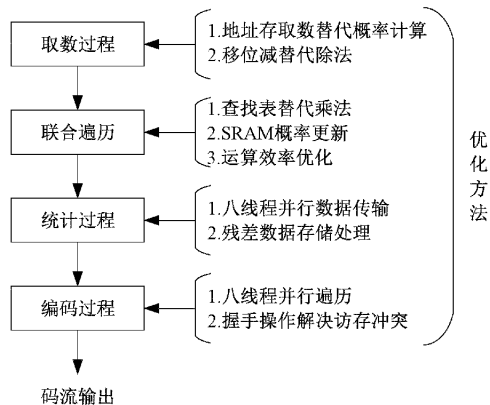


图 3 CABAC 算法映射优化方案

取数过程是根据像素规模标志位选取不同取数方式,

将待编码块进行簇间取数,并存放在对应 PE 的地址中。取数步骤采用八线程并行方式,参考不同块大小标志位,在同一阵列规模上动态改变数据分配方式及编码模式,以满足不同编码规模的高效编码需求,这种数据存储及处理方式能够有效的防止数据冗余,为遍历过程做准备。

联合遍历过程通过 8 个 PE 联合工作,将各 PE 遍历的结果进行组合,构成各元素出现频次的查找表。遍历构建概率查找表的过程是此类编码算法中最为耗时的过程,本文方法有效缩短了概率查找表构建过程中的时间消耗,优化了 CABAC 算法的编码效率,图 4 为查找表构建过程。

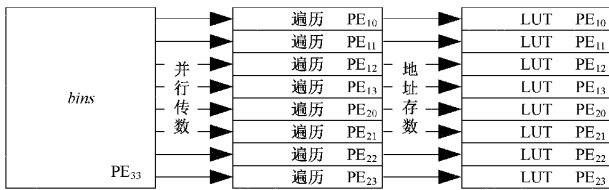


图 4 并行传数、遍历构建查找表

统计过程首先通过移位减法实现的除法操作将每个元素出现频次更新为概率状态 *state*,代表当前概率状态索引符号  $\sigma$  所对应的 *LPS* 出现概率,其中  $\sigma$  是元素出现概率存储的地址编号。用移位减法替代除法能够有效减少运算的资源开销进而提升编码效率,并且在编码输入 *bin* 数据规模越大时,编码效率提升越明显。

通过概率状态更新模型对 *LPS* 进行更新,同时构建概率区间的低位 *low* 以及区间长度 *range* 查找表,用以编码时概率模型更新。采用查找表形式能够用简单的地址存取操作来替代复杂的概率计算,节省时钟周期和资源占用,有效提高编码效率。概率状态转移查找表用以表示每个  $\sigma$  对应的下一个概率状态及概率估计值,将概率估计更新后的概率数据存放在 RAM 中,在下次概率估计更新时对 RAM 进行读取和重新写入,完成概率模型的动态更新。利用查找表进行数据预处理的映射方案如图 5 所示。

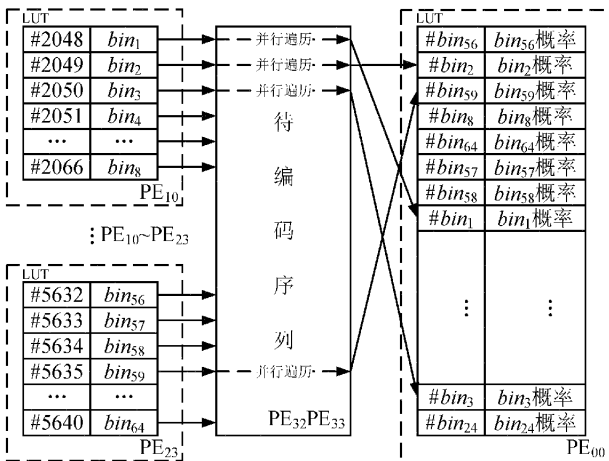


图 5 CABAC 查找表优化映射方案

经过前 3 个步骤的预处理,编码过程的运算和指令操

作步骤被有效简化,降低了单个 PE 在进行编码时的任务量。并行优化后的编码算法能够更好的适应动态重构结构,便于算法的动态重构实现。本文针对各个编码环节的并行优化方法,适用于大多数存在内部并行特性的算法,因此,动态可重构阵列在面向此类算法时具有一定的普适性。

### 3 CABAC 算法动态重构实现

#### 3.1 阵列处理器动态重构机制

图 6 给出了一种典型的动态可重构阵列处理器结构,由主处理器、全局控制器、H 树型配置传输网络、输入输出存储器以及 PE 阵列组成<sup>[13]</sup>。

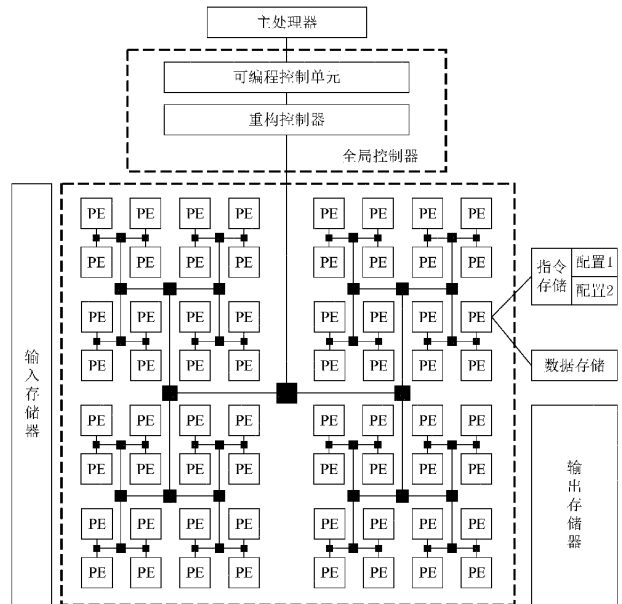


图 6 动态可重构阵列结构

作为动态可重构结构的关键部件,全局控制器完成配置信息管理、阵列状态收集以及数据交互工作。H 树型配置传输网络完成配置信息从上层到阵列的下发,以及阵列执行状态信息到上层的反馈。输入输出存储器用于外部数据存储。PE 阵列包含多个处理元簇,每个簇包含  $4 \times 4$  个处理元(process element, PE)<sup>[14]</sup>。

在算法基于可重构阵列结构的执行过程中,主处理器对阵列结构进行访问,全局控制器接收来自主处理器接口的总线信息,总线信息包括地址信息、标志位信息和指令信息。其中地址信息完成每个 PE 的地址划分,保证不同的 PE 对应不同的地址;标志位信息用于判断执行数据反馈、指令下发或配置调用;指令信息用来确定 PE 所执行的特定指令,将不同配置的 CABAC 算法指令信息初始化在同一 PE 不同的指令存储中。阵列结构能够根据应用需求实现多个算法间的灵活切换,提高应用灵活性和资源利用率。

动态可重构阵列能够根据当前待编码序列的特点动态配置算法所需要的硬件规模。软件重构实现的方法既能满足规模可变的上下文二进制算术编码算法灵活性及速度需

求,又能避免专用器件较高的硬件资源消耗。因此,在可重构阵列处理器上实现 CABAC 算法是解决编码器编码速率低和硬件资源消耗高的可靠方法。

### 3.2 编码过程的可重构实现

编码过程的动态重构由全局控制器执行重构指令,从存储器中读取不同 PE 的配置信息,对配置信息的首地址进行写操作。可编程控制单元通过对控制寄存器进行配置,从而实现对重构控制器的实时配置。重构控制器实现对阵列配置信息的读取、阵列执行状态的实时监测以及阵列资源的自适应调度。由总线信息共同决定每个 PE 在某时刻执行的特定指令,PE 根据配置信息执行相应操作,实现阵列的动态重构过程。通过状态实时监测获取当前重构阵列中各个 PE 的执行状态,并反馈给重构控制器,重构控制器根据实时监测结果实现阵列资源的自适应调度。

由于视频编码算法处理的数据主要是数据块形式,因此,动态重构过程即保证了编码速度同时节约了硬件资源开销。本文提出的 CABAC 的并行化可重构映射方法,采用了 8 线程并行遍历 4 线程并行编码的方式,映射示意图如图 7 所示。

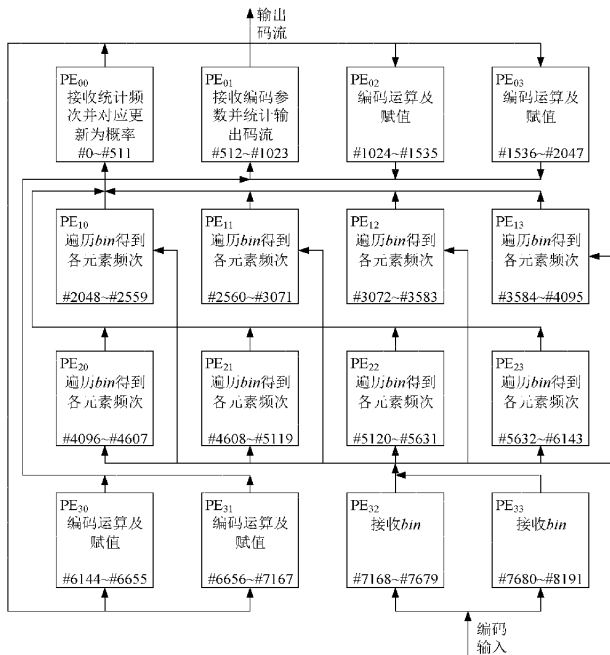


图7 可重构阵列并行映射示意图

当处理器资源有限却有多个算法需要映射时,为了减少硬件资源消耗,动态可重构机制将不同配置的 CABAC 算法指令初始化在同一 PE 的不同指令存储器中。通过配置调用、指令反馈和指令下发等操作来实现 CABAC 算法在可重构阵列处理器上的并行映射,不同 PE 并行化执行算法的不同流程,并根据全局控制器下发的配置信息,对阵列资源进行动态调度,在保证编码速度的同时实现了不同编码规模的 CABAC 算法基于动态可重构阵列的灵活切换。这种基于上下文切换的重构机制能够根据用户需求在

多个算法之间进行切换,最终使用户能够管理和控制资源。

## 4 动态重构仿真及 FPGA 测试

为验证本文提出的可重构实现方法,首先,利用 Matlab 软件对算法可行性进行分析和验证。然后,通过 Questasim 对 CABAC 算法基于可重构阵列处理器的并行映射方案进行仿真,测试映射方案的可行性和编码性能。最后,利用 Vivado 对映射后的阵列结构进行综合实现及结果评估,并基于 Xilinx 公司的 ZYNQ-7000AP 系列 ZC706 评估套件进行 FPGA 测试。

测试结果表明,本文提出的 CABAC 优化算法,基于可重构阵列处理器能够达到 384.13 Mbin/s 的平均编码速度,最高编码速度为 476.84 Mbin/s,工作频率为 150 MHz。能够满足 1 920×1 080@30 fps 视频序列的实时编码要求。

### 4.1 基于 Matlab 的图像重建测试

选取编码结果在 Matlab 中进行图像重建,原图像长度为 65 536,经过本文优化算法压缩编码后码流长度为 2 274,再对编码后的码流进行解压缩操作得到恢复后的图像数据长度为 65 536,数据长度不变。图 8 给出了测试序列的原始图像、二值化后的图像以及对编码后码流解码恢复后的图像对比信息。将编码结果通过 Matlab 解码恢复后的二值图像结果与原二值图像对比可知,二者完全一致,无损压缩特性得以体现。



图8 编解码测试结果

图 9 为编码前、编码后、解码后的图像数据对比。编码后的图像数据码流位数较编码前的图像信息有了明显减少,且解压恢复后的数据基本一致,论文所述 CABAC 编码算法及其优化过程可行。

### 4.2 基于 QuestaSim 的编码仿真分析

为验证本文方法的可行性和编码结果的准确性,基于 QuestaSim 平台搭建可重构阵列结构进行验证。将待编码数据写入到可重构阵列的片外存储器中,再将翻译后的 CABAC 算法二进制指令映射到当前簇的指令 mif 中<sup>[15]</sup>,主处理器根据指令信息对 PE 进行指令下发,完成任务的动态调度,仿真结果如图 10 所示。

可以看出标准测试序列在每个编码周期下能够完成 5 个 bin 的编码任务,编码后的结果传输到 PE00 的 Reg1 寄存器中,实现编码结果的实时更新输出。

图 11 为不同块大小编码周期消耗,其中,不同规模统计了针对标准测试序列的不同块大小基于可重构处理器,进行上下文自适应二进制算术编码时的编码周期消耗。相

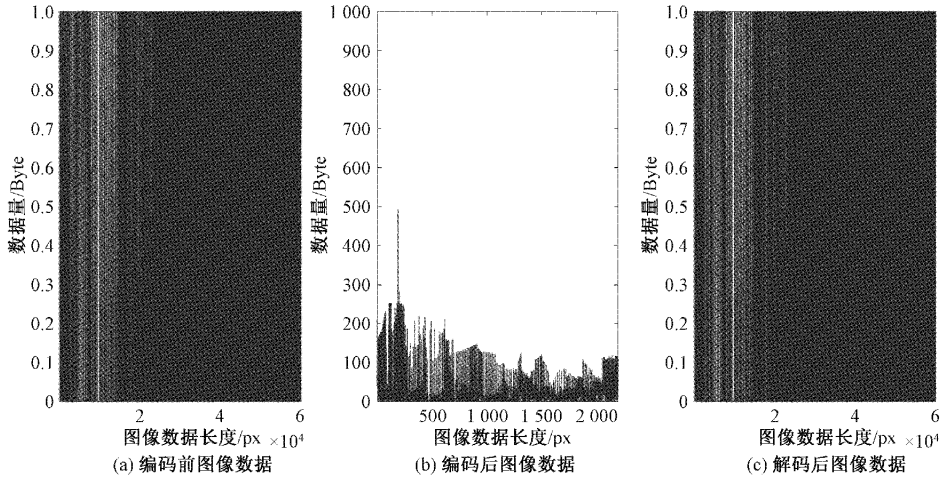


图 9 图像数据测试结果

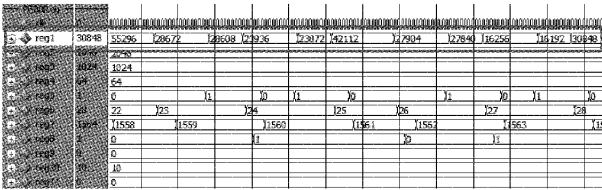


图 10 QuestaSim 仿真波形

同规模统计了在不同的编码速度下,完成 1 024 个 bin 编码的周期消耗。可以看出,当编码规模越大时,完成每个 bin 的编码速度越快,编码效率提升越明显。

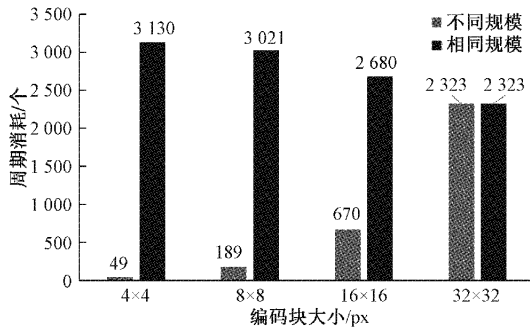


图 11 不同块大小下编码周期消耗统计

### 4.3 FPGA 测试及性能分析

在 Vivado2018.3 中搭建动态可重构阵列处理器平台,将算法指令写入到处理器的片外存储器中,对整个结构进行综合实现,添加 Vivado 的固有 IP 核嵌入式逻辑分析仪 (integrated logic analyzer, ILA) 对编码结果进行板级采样,表 1 给出了单个簇的可重构阵列结构资源使用情况。

表 1 可重构阵列结构资源使用情况

资源	使用数量
LUT	3 013
LUTRAM	407
FF	4 012
BRAM	32.50

综合结果显示,主要资源占用项为查找表与触发器,采用映射方法实现的 CABAC 算法相较于专用器件的资源占用更少,得益于可重构阵列结构的灵活性优势,使编码过程具备了专用器件进行复杂编码运算时所具有的速度优势,表 2 为与 CABAC 算法并行化专用硬件结构对比结果。

表 2 CABAC 编码算法映射对比

结构	硬件工艺/ 验证平台	视频 标准	支持块 大小	可否 重构	硬件资源	编码速度/ (Mbin/s)	吞吐量	最高主频/ MHz
文献[6]	Arria10 ME	VVC	4~64	否	9723(LUT)+14368(Reg)	363	1 920×1 080@48 fps	165
文献[7]	Virtex7	VVC	8~32	否	9313(LUT)+3515(FF)	451.7	1 920×1 080@47 fps	227
文献[16]	XCVU9P	VVC	4~64	是	2382(LUT)+3106(FF)	357	1 920×1 080@34 fps	119
文献[8]	xc6slx92tqg144	HEVC	4~32	否	4172(LUT)	89×4	1 920×1 080@30 fps	89
文献[9]	Arria II GX	HEVC	8~32	否	3779(LUT)+1304(Reg)	400	NA	115
本文	Zynq-ZC706	VVC	4~64	是	3013(LUT)+4012(FF)	384.13	1 920×1 080@30 fps	150

与文献[6]的专用硬件结构设计相比,本文的动态重构结构在支持大小为 $4 \times 4 \sim 64 \times 64$ 的编码块的同时编码速度提升了5.47%,虽然数据吞吐量和最高主频略低,但能够有效避免其较高的硬件资源开销。文献[7]的专用硬件结构的编码性能较本文更优,平均编码速度提高了17.63%,最高主频提高了51.33%,但输入的块大小限制了编码结构的灵活性,本文的硬件资源占用率降低了一倍以上,体现出基于软件重构方式实现编码器的面积优势。文献[16]采用了可重构结构设计,结构灵活资源占用率低,本文与其编码部分相比硬件资源占用率提高了20.94%,编码速度提高了7.03%,最高主频提高了20.66%。文献[8]的专用硬件结构有着较少的硬件资源消耗,本文方法的结构设计与之相比编码速度提升了7.29%,最高主频提高了68.53%,且资源占用率降低了27.78%。文献[9]的专用硬件结构与本文相比,数据吞吐量较大但对处理的最小规模有一定要求;在资源利用率有效降低的情况下,本文的最高主频提高了30.43%,最高编码速度提升了19.11%。

## 5 结 论

针对H.266/VVC视频编码标准中的自适应二进制算术编码器编码速度慢、资源开销大问题,论文基于动态可重构阵列处理器提出了CABAC编码的软件重构实现方法,在可重构阵列处理器上实现了VVC中CABAC算法常规编码模式下的并行映射。实验结果表明,在保证 $1920 \times 1080@30$  fps的编码吞吐量前提下,本文方法能够根据不同规模编码序列实现算法结构的灵活切换,对于CABAC编码算法的可重构映射实现能够达到384.13 Mbin/s的平均编码效率,最高编码效率达到476.84 Mbin/s。总之,本文方法不仅可以满足通用视频编码标准下的实时编码要求,也避免了专用硬件较高的资源开销,并且能够根据编码任务的不同需求,对阵列结构的规模和执行模式进行重新配置,能够广泛适用于各种数据密集型 and 运算密集型算法,是具有潜在并行性算法高性能硬件实现的有效途径。

## 参考文献

- [1] PAKDAMAN F, ADELIMANESH M A, GABBOUJ M, et al. Complexity analysis of next-generation HEVC encoding and decoding[C]. ICIP 2020: 2020 IEEE International Conference on Image Processing, Abu Dhabi: IEEE, 2020: 3134-3138.
- [2] WEI S J, LI Z S, ZHU J F, et al. Reconfigurable computing: Toward software defined chips[J]. SCIENTIA SINICA Informationis, 2020, 50(9): 1407-1426.
- [3] 石硕,王瑞雪,李慧,等. LDPC码的多路并行编码器实现[J]. 电子测量与仪器学报, 2021, 247(7): 83-89.
- [4] 张珊珊,赵建华. 基于压缩感知的重建算法仿真分析[J]. 国外电子测量技术, 2019, 38(10): 44-48.

- [5] 于晓升,许茗,王莹,等. 基于卷积变分自编码器的异常事件检测方法[J]. 仪器仪表学报, 2021, 42(5): 151-158.
- [6] FARHAT I, HAMIDOUCHE W, GRILL A, et al. Lightweight hardware implementation of VVC transform block for ASIC decoder[C]. ICASSP 2020: 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing, Barcelona: IEEE, 2020: 1663-1667.
- [7] AZGIN H, KALALI E, HAMZAOGLU I. An approximate versatile video coding fractional interpolation hardware[C]. ICCE 2020: 2020 IEEE International Conference on Consumer Electronics, Las Vegas: IEEE, 2020: 1-4.
- [8] SHU W, KUANG Z, ZHANG G, et al. Fast hardware implementation of renormalization for context-based adaptive binary arithmetic coding[C]. ICIASE 2019: 2019 IEEE International Conference of Intelligent Applied Systems on Engineering, Fuzhou: IEEE, 2019: 313-316.
- [9] ZHANG Y, LU C. A highly parallel hardware architecture of table-based CABAC bit rate estimator in an HEVC intra encoder[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(5): 1544-1558.
- [10] KARWOWSKI D. Precise probability estimation of symbols in VVC CABAC entropy encoder[J]. IEEE Access, 2021, 9(1): 65361-65368.
- [11] KARWOWSKI D. Precise estimation of probabilities in CABAC using the cauchy optimization method[J]. IEEE Access, 2020, 8(1): 32088-32099.
- [12] 施金诚,杨静. 基于深度学习的VVC快速帧内模式决策研究[J]. 电子测量技术, 2022, 383(3): 104-111.
- [13] 蒋林,武鑫,崔继兴,等. HEVC运动估计中SAD算法的动态可重构实现[J]. 北京邮电大学学报, 2018, 41(4): 37-43.
- [14] 杨坤,蒋林,谢晓燕,等. HEVC中率失真优化算法的动态可重构实现[J]. 计算机工程与科学, 2021, 43(2): 354-361.
- [15] 蒋林,贺飞龙,山蕊,等. 可重构视频阵列处理器测试平台设计与实现[J]. 系统仿真学报, 2020, 32(5): 792-800.
- [16] CAI Y, FAN Y, HUANG L, et al. A fast CABAC hardware design for accelerating the rate estimation in HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 4(1): 2385-2395.

## 作者简介

刘尧,硕士研究生,主要研究方向为专用集成电路设计、计算机体系结构。

E-mail:lyao361@163.com

蒋林(通信作者),工学博士,教授,主要研究方向为专用集成电路设计、计算机体系结构、计算机图形图像处理。

E-mail:jianglin@xust.edu.cn

李远成,工学博士,讲师,主要研究方向为计算机体系结构、并行计算、人工智能、CCF会员(41522M)。

E-mail:yuanch\_li@126.com

山蕊,工学博士,副教授,主要研究方向为专用集成电路设计、计算机体系结构。

E-mail:shanrui0112@163.com