

Efficient Macroblock Pipeline Structure in High Definition AVS Video Encoder VLSI Architecture

Hai bing Yin^{1,2}
The Institute of Digital Media¹
Peking University
Beijing, China

Hong gang Qi¹, Huizhu Jia¹, Don Xie¹, Wen Gao¹
Zhejiang Provincial Key Laboratory of Information Network
Technology, Zhejiang University²
Hangzhou, China

Abstract—In traditional four-stage pipeline structures for H.264 video encoder hardware implementation, rate distortion optimization (RDO) based mode decision was turned off, and dual-port or ping-pong on-chip search window SRAM was used to achieve data reuse between the integer and fractional pixel motion estimation. To support RDO based mode decision for efficient high definition AVS video coding implementation, we propose an improved four-stage MB pipeline structure. Also on-chip buffer structure is optimized to achieve the balance between circuit consumption and coding performance. The Jizhun profile AVS video encoder is successfully mapped into hardware implementation with the proposed pipeline structure with small performance degradation.

I. INTRODUCTION

AVS-P2 is the video part of the China audio video coding standard (AVS) that achieves good trade-off between performance and complexity. It had been formally accepted as the national standard of china. The industrialization process is being on led by the AVS industry alliance.

Currently dedicated AVS video encoder chip is still vacant. Although the complexity of AVS is lower than H.264, real-time high definition (HD) AVS video coding is still a huge challenge. Implementation on ASIC or programmable VLSI (FPGA) is highly desired.

Several works [1] [2] were reported on 720P or 1080P H.264 video encoder VLSI implementation, which is very challenging due to the high throughput. Parallel pipelining technology was widely used in VLSI implementation to achieve the data processing throughput. Thus, macroblock (MB) level pipeline structure is important for the whole encoder architecture design and implementation.

In baseline profile H.264 video encoder architectures [1]-[3], four or three stage MB pipeline structure was adopted. B frame was not supported and only one reference frame was used with relatively small search window (SW) for motion estimation. Also, simplified mode decision is used based on the SATD (Sum of absolute transformed difference) criterion.

Differing from H.264 baseline profile, bidirectional B frame is supported in the basic AVS Jizhun profile. Thus, the resulting on-chip SW buffer size and external SDRAM access

bandwidth are both doubled. Also, some coding tools are simplified in AVS, such as less MB partition and intra prediction modes. If simplified mode decision is also used in AVS video encoder with small motion SW, the resulting coding performance degradation will be obvious. Thus, RDO based mode decision and motion estimation with large SW are adopted in our AVS video encoder. Conventional four-stage pipeline structure is highly challenged if it is applied in AVS video encoder, and it should be optimized according to the AVS algorithm. Moreover, on-chip buffer is another important problem to be considered in MB pipeline structure design.

In this paper, we focus on pipeline structure optimization for HD hardware AVS video encoder. The AVS introduction and problem analysis are given in section II. The proposed improved MB pipeline structure is given in section III. Simulation results and conclusion are drawn in section IV.

II. PROBLEM STATEMENT AND ANALYSIS

The block diagram of AVS video encoder is given in Fig.1. Similar to H.264, MB is the basic processing unit in AVS, and the major modules include motion estimation (ME), motion compensation (MC), intra prediction (IP), mode decision (MD), residue coding loop (DCT/Q/IQ/IDCT), deblocking filter (DF) and entropy coding (EC). AVS Jizhun profile is similar with main profile H.264, in which B frame with bidirectional motion estimation is supported.

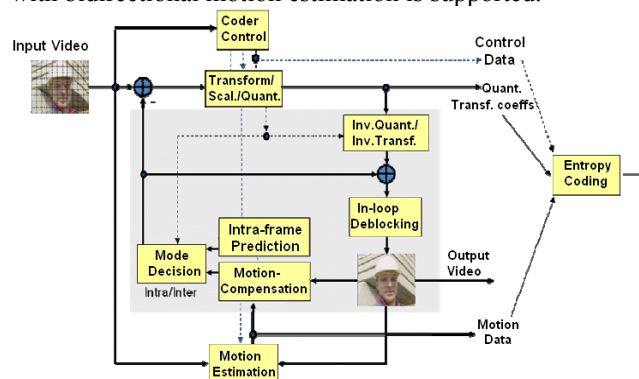


Fig.1 The Block diagram of AVS video encoder.

Our design target is an hardware video encoder for AVS Jizhun profile with ± 128 horizontal and ± 96 vertical integer

This work was supported in part by NSFC 60802025, 60833103, and the open project of Zhejiang Provincial Key Laboratory of Information Network Technology (200815).

pixel SW. This target is challenged by the high throughput due to HD image and bidirectional ME. MB level pipelining is inevitable to achieve the desired throughput.

A. Challenge Analysis of Pipeline Structure

MB pipeline structure suffers from some challenges including data dependency, high throughput, and the balance between circuit consumption and coding performance.

Data dependencies in video coding algorithms disturb the normal pipeline rhythm [1]. At the MB level, integer pixel ME (IME), fractional ME (FME), MD and IP, EC and DF are processed in turn. At the block level, one block IP can't initiate until its left, up blocks have been reconstructed; In the residue coding loop, DCT, Q, IQ, and IDCT are processed in turn. The motion vector (MV) prediction is context-dependent with the up, left, and up right blocks. These dependencies are harmful for normal pipeline rhythm.

High throughput is another challenge [2]. In HD video coding, there are so many MBs to be processed within a second. This high throughput results in high clock frequency or hardware parallelism. Also, RDO based MD is challenged by high throughput due to multiple coding modes to be selected. Another throughput challenge is the external SDRAM memory bandwidth. IME confront high circuit parallelism and high access burden between on-chip and off-chip memories in HD cases.

Achieving balance among circuit consumption and coding performance is the third challenge [3]. Some coding tools ill-suited for hardware implementation contribute to trivial performance improvement in HD AVS video encoder. It is crucial to tailor the algorithms in a hardware friendly configuration for MB pipeline structure optimization.

B. Typical MB Pipeline Structure in Previous Works

Typical four-stage H.264 video encoder MB pipeline structure is shown in Fig.2 [1]. They are IME, FME, IP/MD, EC/DF. This typical pipeline structure solves the problems of throughput and data dependency with SATD based simplified MD. IP and MD are both at the third stage, and the residue coding loop is employed at this stage for MB reconstruction. There is no data dependency between EC and DF, and they are processed in parallel at the same stage. Three-stage pipeline structure was adopted in [2], in which FME and IP are combined in the same stage with algorithm simplification. If there are no throughput conflict between FME, IP and MD, this three-stage pipeline structure is cost-efficient and can save shared buffer between IP and FME.

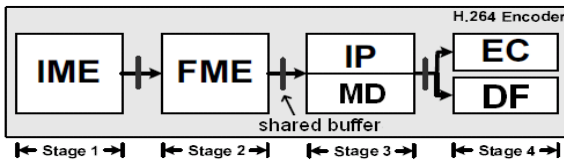


Fig.2. Typical four-stage MB pipeline structure in H.264 video encoder.

C. Consideration for MB Pipeline in AVS Video Encoder

In two typical MB pipeline structures, simplified SATD based MD is employed. This simplification results in obvious

performance degradation. Also, only P frame with one reference frame is supported, or relatively small SW is used for IME [1], or simplified FME algorithm is used in three-stage pipeline structure [2]. Also, three-stage pipeline is ill-suited for HD video encoder if RDO based MD is used due to the throughput burden. Thus, four-stage pipeline structure will be used in our AVS video encoder. Algorithm and architecture optimization is necessary for RDO based on MD support.

First, in the proposed HD AVS video encoder, large SW 256x192 is targeted. Two reference frames need large on-chip SW SRAM. Double-buffered ping-pong or dual-port SMAM was used in conventional pipeline structures to achieve data share between IME and FME. However, these two kinds of SW SRAM consumption in AVS video encoder will be too high due to two reference frames. Thus, single port SW SRAM is highly desired, and efficient SRAM share between IME and FME is still necessary for on-chip SRAM saving. Level C+ data reuse scheme [2] is also adopted for SDRAM bandwidth burden alleviation in this work.

Second, less modes in AVS compared with H.264 make hardware implementation for RDO based MD more possible. In this work, genuine RDO based MD is targeted. The residue coding loop and the EC loop are both embedded at the MD stage for distortion and bit rate estimation. However, genuine coding for the final selected mode also needs to perform the residue coding and EC loops. There is redundant hardware if the residue coding and EC loop are both adopted at the IP/MD and EC/DF stages. Thus, circuit share between the third and the fourth stages is necessary here.

Third, hardware oriented algorithm simplification is desired to coordinate with MB pipeline structure. On the one hand, data dependency problem should be solved by algorithm simplification. On the other hand, algorithm modification is necessary to trade off hardware complexity and coding performance. Also, it is an important problem for hardware reuse between fractional pixel interpolation of the skip/direct, symmetric mode, and normal forward or backward FME.

III. PROPOSED MACROBLOCK PIPELINE STRUCTURE

The proposed pipeline structure is shown in Fig.3 with algorithm simplification and architecture customization.

A. Improved Four-stage Pipeline Structure

In Fig.3, the improved four pipeline stages are IME, FME, IP/MD, and bitstream generation (BG) and DF. This structure modification is mainly derived by RDO based MD., which is selected by minimizing rate distortion cost function $RDcost = D + \lambda \times R$. The residue coding loop is employed for distortion (D) estimation, and the EC loop (zigzag scanning, run-length coding, Exp-Golomb coding) is employed for bit rate (R) calculation. Thus, RDcost estimation computation for all modes is very high.

In conventional H.264 pipeline structure, inter modes are selected at the FME stage according to SATD, and intra modes are selected also according to SATD at the IP stage with residue coding loop embedded for pixel reconstruction and final mode decision between intra/inter modes, but not for selection of all candidate modes.

In conventional four-stage pipeline structure, the EC loop is not included at the IP/MD stage. Thus, the EC loop for the final coding mode is arranged at the EC/DF stage. However, in RDO based MD pipeline structure, both the residue coding loop and the EC loop are both embedded at the IP/MD stage for RDcost estimation and MD. Hardware parallelism is necessary in the residue coding and the EC loop engines at the IP/MD stage to achieve the desired throughput. To avoid redundant hardware consumption, the residue coding and the EC loop engines are shared between the third and the fourth stages. The EC loop at the third stage provides the CodeNum data for BG at the fourth stage. With the CodeNum field, BG only needs to perform simple exp-Golomb coding and syntax-compliant bitstream generation. Thus, only DF and BG are needed at the fourth stage.

As shown in Fig.3, the forward (Forw.) and the backward (Back.) IME/FME engines are adopted in the proposed architecture in parallel to perform forward and backward ME for B frames. In P frames, two reference frames are used and searched respectively by two parallel IME and FME engines.

B. On-chip Buffer Structure Optimization

On-chip buffer structure is very important because that on-chip SRAM generally consumes more than 50% circuit gate budget. Especially in HD AVS video encoder with B frame, the SW buffer is the largest SRAM consumer. In conventional four-stage H.264 architecture, only one reference frame is used. In order to implement SW buffer data share between IME and FME, dual-port SRAM or double-buffered single SRAM are used. If this SW buffer is adopted in this work, the SW SRAM consumption will be too high and unacceptable.

To reduce the SW SRAM consumption, we have proposed an intelligent SW buffer structure for IME and FME data share with single port SRAM [4]. According to previous research, VBSME can be done within a local small SW centered about an appropriate center MV (*mvp*) instead of the whole SW if we can predict *mvp* accurately enough. This simplification result in negligible performance degradation.

As shown in Fig.3, the forward and backward SW reference pixels are stored in single-port *Forw. Luma Ref. Pels SRAMs* and *Back. Luma Ref. Pels SRAMs*. They are sixteen SRAMs, and the whole SW are interlaced and stored into them by two level 4:1 downsampling. Data format translation and buffering between SDRAM and these two SW buffers is achieved by *Forw. Luma Ref. Reg Array* and *Back. Luma Ref. Reg Array*, whose size is very small. Multiresolution IME predict the center MV (*mvp*) first, then variable block size ME (VBSME) is performed and the local small luma SW is transferred simultaneously into the dual-port *Local Luma Ref. Pels SRAMs*, by which efficient data share between IME and FME is achieved. Using this buffer share mechanism, almost 50% on-chip SW buffer can be saved without any SDRAM bandwidth access increase.

The chrominance (chrom) components do not join IME and FME, thus it is unnecessary to load the whole chrom SW into on-chip buffer. According to *mvp*, we can only load the corresponding local small chrom SW, i.e. *Local Forw. Chrom Ref. Pels SRAM* and *Local Back. Chrom Ref. Pels SRAM*. Similarly, the *Forw. Chrom Reg. Array* and *Back. Chrom Reg. Array* are employed to perform format transform and buffering. Thus, this local SW buffer can save 80% chrom SW SRAM consumption compared with the unoptimized case.

The 1/4 pixel interpolation versions of the displaced blocks of all possible inter mode are buffered in the *Luma Pred. Pels SRAMs* (part I and II) and *Chom Pred. Pels SRAM* (part I and II) to implement data share between FME and IP/MD stages.

To achieve circuit reuse of the residue coding and the EC loops between IP/MD and BG/DF stages, the *MB CodeNum SRAM* is employed to store the CodeNum fields of all coefficients in the blocks of the selected optimal mode. Thus, bitstream can be easily generated at the following BG stage according to the CodeNum using Golomb exp-coding, and the coded bitstream is buffered in the *Bitstream SRAM* to wait for SDRAM bus transactions.

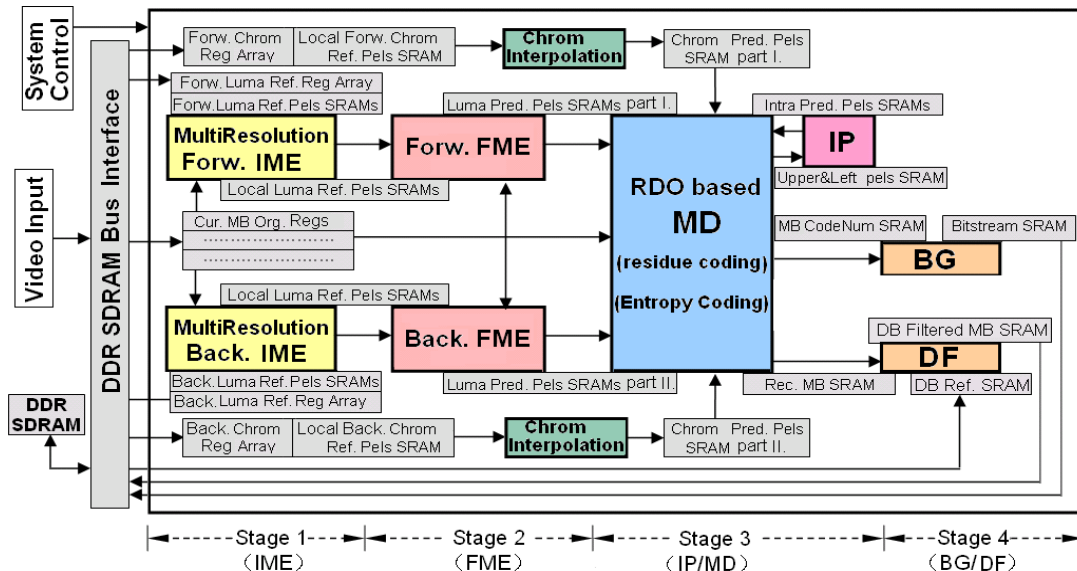


Fig. 3. The proposed pipeline structure and system architecture for HD AVS video encoder hardware implementation.

C. Hardware Oriented Algorithm Simplification

Multiresolution IME with three hierarchical levels is adopted. Full search is done at two coarse levels sequentially using 16 parallel ME processing element (PE) arrays to achieve searching 16 candidate MVs in each cycle at the coarsest level, and 4 candidate MVs in each cycle at the middle level. With this PE array structure, we can achieve the throughput of SW 256x192, and efficient PE array circuit share at three levels is achieved. After two coarse level full searches, the obtained *mvp* is used as the center for full search VBSME within the local small SW 32x24 at the finest level. 16 way parallel PE array collaborate to achieve one candidate MV in each cycle at the finest level.

The simplified MV prediction algorithm similar with [1] is adopted to break the data dependency. Also, bidirectional (symmetric) prediction is used in AVS standard, symmetric mode is only used in FME in our work to alleviate the burden of fractional pixel interpolation computation and to avoid disturbing normal FME pipeline rhythm. The forward FME result is used as the FME refinement center for symmetric mode. With this simplification, the extra interpolation computation needed is acceptable and has no conflict with FME pipeline rhythm.

In AVS, there are five luma modes and four chrom modes in intra prediction. There are five inter modes including skip, 16x16, 16x8, 8x16, and 8x8 in P frames. The inter prediction modes of B frames are more complex. An inter prediction mode of B frame is related with two factors. One is prediction direction (forward, backward, and bidirectional). Another factor is the MB partition mode such as 16x16, 16x8, 8x16, and 8x8. The two factor combinations result in abundant inter modes. If these two factors are selected by RDO based MD, there may be evenly seventy candidate prediction modes with intra modes considered. As a result, the throughput is still too high. Necessary simplification is highly desired.

In this work, genuine RDO based MD is adopted for intra mode selection in I frames, while SAD based MD is used for intra mode selection in P, B frames. Two factors in MB inter prediction modes in P and B frames are separately selected. Temporal prediction direction is pre-selected at the FME stage, and the MB partition mode selection is done by genuine RDO based MD. With two simplified measures, candidate modes and MD hardware parallelism are largely reduced.

IV. SIMULATION AND RESULT ANALYSIS

Hardware oriented C model is developed based on the AVS reference code RM52J. Four sequences ‘‘Sailormen’’, ‘‘City’’, ‘‘Night’’, and ‘‘Spincalendar’’ of 720P format at 30Hz are used for simulation. The later two sequences have complex motion and are selected for rigorous performance evaluation. IPBBPBB format with GOP length 15 are used. SW 256x192 is used for 1/4 pixel VSBME. All inter/intra modes and RDO based mode decision are supported.

We select the algorithm with full search ME and RDO based MD without any simplification as the anchor (FS + RDO). The rate distortion curves of the proposed C model with algorithm simplification and the perfect anchor are given in Fig.4. According to the results, the rate distortion performance degradation due to the algorithm simplification is relatively small and acceptable.

We use Verilog-HDL to implement the hardware design with function verification based on Virtex5 FPGA ASIC development system. The Synplify Pro is used as synthesis tools for FPGA case. Efficient hardware architecture results in reasonable circuit and SRAM consumption. The detailed parameters are given in table I, and further optimization is being on for further gate saving.

The major features in main profile H.264 and Jizhun profile AVS are very similar. Thus, the proposed work is also well suited for main profile H.264 video encoder.

TABLE I
PERFORMANCE OF THE PROPOSED MD ARCHITECTURE

Item	Parameters
Capacity	108Mhz (720P 30fps), 220Mhz (1080P 30fps)
SRAM	162Kbytes equivalent single port SRAM
Modes	RDO based MD for all intra and inter modes
Search range	256x192, bidirectional B frame supported
Gate count	81% slice of Xilinx FPGA V5LX330 1380k equivalent ASIC Gates

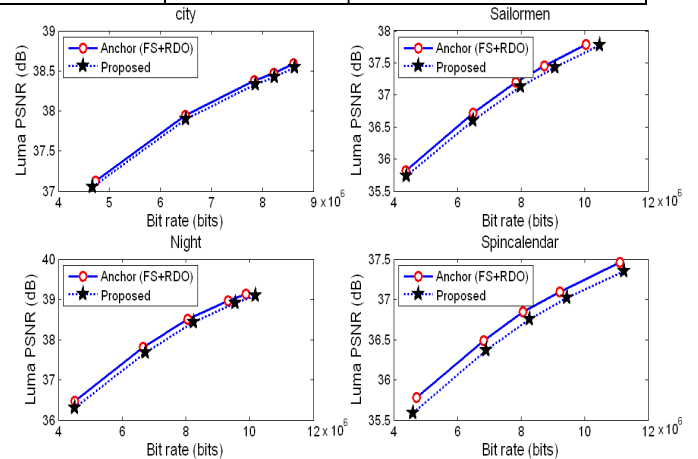


Fig.4. The rate distortion curves of various standard sequences of the proposed software and the perfect anchor with full search and RDO.

REFERENCES

- [1] T.C. Chen and et al, "Analysis and Architecture Design of an HDTV 720p 30 Frames/s H.264/AVC Encoder," IEEE Trans. Cir. Syst. Video Tech., vol. 16, no. 6, pp. 673-688, June 2006.
- [2] Zhenyu Liu, Yang Song, Satoshi Goto etc. HDTV 1080P H.264/AVC Encoder Chip Design and Performance Analysis, IEEE Journal of Solid-state Circuits, Vol.44, no.2, Feb,2009.
- [3] Tung-Chien Chen, Yu-Wen Huang, and Liang-Gee Chen, "Analysis and design of macroblock pipelining for H.264/AVC VLSI architecture", IEEE ICASS, Hong Kong, China, April, 2004.
- [4] Hai bing Yin and et al, VLSI Friendly ME Search Window Buffer Structure Optimization and Algorithm Verification for High Definition H.264/AVS Video Encoder, IEEE Conference on Multimedia and Expo, Cancun, Mexico, June 28-July 3, 2009.