THE IMPACT OF ENCODING ALGORITHMS ON MPEG VLSI IMPLEMENTATION

Sheu-Chih Cheng and Hsueh-Ming Hang

Department of Electronics Engineering
National Chiao-Tung University,
Hsinchu, 300 Taiwan, ROC

ABSTRACT

The goal of this paper is to study the impact of MPEG encoding algorithms from a system-level design viewpoint. An area-time estimation tool is developed to extract the timing requirement and silicon area for various combinations of hardware modules and algorithms. After complemented the design of several modules in an MPEG encoder, we found that the motion estimation and rate control modules consume most part of the silicon area in the encoding chip. We also evaluated the entire chip area of a few cases for two picture formats. The methodology and results presented here should provide useful guidelines in selecting an appropriate MPEG encoding algorithm for VLSI design.

1. INTRODUCTION

The goal of system-level VLSI design is selecting or designing proper algorithms and architectures to achieve a low chip fabrication cost and good coding performance. It is often observed that a properly chosen algorithm can significantly reduce the hardware cost. However, it is unlikely for a system-level designer to implement all possible candidate algorithms and calculate their exact silicon area and performance. We need a chip area estimation tool that can provide a rough silicon area estimate at algorithm level. On the other hand, silicon area is not the only cost measure of an industrial product such as an MPEG encoder. There are tradeoffs among various hardware costs and performance factors in designing an MPEG VLSI chip. Without detailed analysis on the encoder modules, it is difficult, relying only on the heuristics, to choose a well-designed algorithm/architecture that ensures a good performance and keep a minimum cost.

Although the MPEG decoder is well-defined by the ISO standards, the MPEG encoder is not completely specified in the standard. Therefore, in addition to the selection of hardware architectures, we chose *different* encoding algorithms for each module in the encoder. The goal of this paper is to show how the MPEG coding algorithms affect the VLSI design. Hence, it is not our intent to devise new or better architectures for a specific component in the MPEG encoder, but rather we like to demonstrate how different coding algorithms would affect the chip area, picture quality, peripheral devices, etc. To achieve this goal, one of our contributions is to propose a method to estimate the timing requirement and chip area for a particular choice of architecture and algorithm. The methodology proposed here in evaluating the algorithm efficiency is quite general, and thus the results found in this study should also be useful to the VLSI design of other types of complicated systems.

2. AREA-TIME ESTIMATION MODEL

An area-time estimation method is developed to extract the area-time information for various algorithms and architectures combinations in this paper. This tool can estimate the silicon area and I/O bandwidth of an MPEG module for a given set of design parameters including picture format, search range, circuit architectures, and various algorithms. A simplified flowchart of this tool is shown in Fig. 1. The results of the area-time estimate can identify the critical modules in designing the MPEG coder from the system-level design viewpoint.

The area-time estimation tool is composed of three major units: clock speed estimation, structure description, and area estimation. In the clock speed estimation unit, the timing requirement of each hardware component is calculated based on the parameters of picture format. The bandwidth requirement of the external RAM is also specified by this unit. The structure of MPEG encoder described in the second unit (structure description) is the core of this tool. This unit converts the system functional descriptions into structural descriptions using the *Architecture Library*. The architecture library contains pre-defined architectures of various subsystems in an MPEG coder, e.g., six motion estimation modules, two variable wordlength coding (VLC) modules, etc. A compression algorithm is transformed into logical level hardware component description using these existing structures. The resultant structure description and timing

0-7803-4455-3/98/$10.00 © 1998 IEEE

requirement are sent to the silicon area estimation unit to estimate the chip area.

In order to do so, a *Component Library* is constructed, which contains a number of key IC components such as adders and multipliers. They are pre-designed using Synopsys and a 0.6 μm single poly double metal (SPDM) standard cell library. Different design parameters such as picture format and architecture may lead to different clock rate requirement and word length. These requirements are used to pick up proper versions of components. Then, the total chip area is the sum of all the component areas. In case that a specific clock rate or word length is not included in the pre-designed component library, interpolation function is performed to obtain a rough estimate of component silicon area.

3. MPEG ENCODER STRUCTURE

The block diagram of our MPEG encoder structure is shown in Fig. 2. Specifically, the memory management module acts as an interface between the processing modules and the external RAM, because the memory access bandwidth is a bottleneck in the encoder chip. The RISC controller performs the functions of bit allocation, rate control, and the other coding parameters adjustment. The operations of DCT/IDCT and VLC modules are defined by the standards. Hence, algorithm variations will by and large appear on the Motion Estimation and the Coding Control (implemented by a RISC processor) modules. Since the impact of motion estimation algorithms and rate control algorithms on VLSI design have been reported in [1] and [2], respectively, this paper focuses on the overall MPEG encoding chip design.

Due to limited space, the processing modules that use existing architectures can only be briefly summarized. The *Motion Estimation* module performs the function of block-matching motion estimator. Many motion estimation algorithms can be implemented using this basic systolic array structure but their chip areas are rather different. To simplify the comparison, only the exhaustive search (ME_{full}) and the three step search (ME_{tss}) which represents fast searching algorithm are included here.

The frame memory processor module (FMP) is used to perform the function of motion compensation at encoder. It imitates the job of decoder to reconstruct the reference frame. The discrete cosine transform (DCT) has been widely recognized as an efficient technique for image compression and is adopted in many image standards. The $L \times L$ size 2D-DCT is decomposed into two length L 1D-DCT. A direct implementation of an L length 1D-DCT requires $2L^3$ operations. Several architectures have been proposed to reduce its computational complexity. For example, one popular architecture uses the systolic array [3], whereas another approach uses the distributed architecture [4]. Both approaches are included in the *Architecture Library*.

In MPEG encoding system, the quantization module (Quan) is essentially a division operation. There are two division operations included in this module. One way to implement the divisional operation is to use a shift-add array. However, we can also implement it by a multiplier and an approximation of the reciprocal of the divisor stored in ROM. Many structures of VLC codec have been suggested to support various coding applications. They can roughly be classified into two types: with or without the barrel shifter. For example, in Lei's structure [5], the parallel concatenation of codewords is done by three barrel shifters, which means it can send a codeword in one clock cycle. On the other hand, Prasanna's structure [6] does not require the barrel shifter.

The *Rate Control* module (RCTL) is an essential element in a video coding system in order to transmit the coded bit-stream over a constant bit-rate channel. To our knowledge, the existing rate control algorithms can be classified into three groups according to their quantizer selection operation, i.e., buffer-feedback method ($RCTL_{TM5}$), budget planning method ($RCTL_{BP}$), and optimal bit allocation method in rate-distortion sense ($RCTL_{OPT}$). These three algorithms have been analyzed and compared in [2] based on the criteria of silicon area, size of internal buffer, and image quality. Memory management module (MMM or MMU) is also an important part in the encoder chip, because a serious bottleneck exists in the data acquisition from the external DRAM.

4. EXPERIMENTAL RESULTS

4.1. Area Estimation

As an example, the MPEG encoder chip area for two picture formats and two motion vector search ranges is shown in Table 1. In this table, the meaning of *Quan*, FMP, and MMU are described in Sec. 3. They are the modules for quantizer, frame memory processor, and memory management. The items of DCT_1 and DCT_2 are the DCT module using the distributed architecture [4] and systolic array architecture [3], respectively. VLC_1 and VLC_2 are the VLC module with and without the barrel shifter [5][6].

One may notice that the area of the exhaustive search motion estimation (ME_{full}) and the optimal bit allocation algorithms ($RCTL_{OPT}$) are approximately 10 times larger than the other units and thus they become the dominate factor. For larger sizes of pictures, three-step search is more practical. This table also shows that the budget planning rate control algorithm has a significant advantage in hardware cost while maintains a comparable rate-distortion performance. The architecture selection of certain modules has an impact on chip area such as DCT, while the others do

not make much difference such as VLC. Since our area estimation model cannot be very precise, the actual silicon area would be somewhat different from our estimates.

In additional to ordinary computational complexity of compression algorithm, the on-chip buffer size is very critical in VLSI realization. However, due to asynchronous pipeline modules in the MPEG coding chip, internal buffers are necessary to store the temporary data generated by each processing unit. If the overall internal buffer is very large, it takes a lot of chip area. Hence, we also evaluate the internal buffer size in each module. The size of on-chip memory depends on the chosen architecture design. In an MPEG coding system, because the data order in several processing modules are rather different, these modules have to wait for all the data being collected before processing. For example, one complete block of data (64 image pixels) are needed in the DCT module to calculate the first coefficient. Hence, the minimum size of the internal buffer is one block of data. We use this intrinsic property to estimate the minimum internal buffer size in each module. We find that both the *DCT* and the *SCAN* operations need a block of data buffer. In the *Quan* and *VLC* modules, data can be sequentially in and out. Hence, only one single coefficient buffer is needed. The sizes of internal buffers are listed in Table 2. In this table, we neglect the small internal buffer size in the modules that needs only one coefficient buffer. The internal buffers in *Motion Estimation* and *Rate Control* modules have been analyzed in [1][2]. The total area of the internal buffers is about 10.2 mm^2 under the assumptions that the external RAM bus width is 60 bits, the budget planning rate control algorithm is in use, and the three-step search of search range 15 is adopted with type B buffer. From Table 2, it is clear that internal buffers are also critical in MPEG coding chip design.

4.2. Picture Quality

In an MPEG coding algorithm, motion estimation and rate control are two crucial elements in determining the picture quality and compression efficiency. Without sacrificing the picture quality, the three-step search algorithm can be used to reduce hardware complexity. The motion estimation search range is 47 for P-pictures and 15 for B-pictures. Several sequences have been tested. Two of them are reported here. Figure 3 shows the PSNR performance for the CCIR *football* image sequence with a target coding rate of $5Mbits/s$. Figure 4 shows the PSNR performance in the *Gaussian* picture sequence, which is a salesman sequence surrounded by Gaussian noise. If we only count the center of pictures in PSNR calculation, it is clear that the budget planning algorithm outperforms all the other algorithms. However, the optimal bit allocation algorithm has the best overall PSNR performance when the surrounding noise background are included. In general, the budget planning algorithm has a much lower hardware cost and a slightly lower objective PSNR, but its subjective quality is as good as the other algorithms if not better.

5. SUMMARY

We found that choosing an efficient motion estimation algorithm and a rate control algorithm is very critical in designing an MPEG VLSI encoder. In addition to the picture quality, the hardware implementation issue is essential for industrial products. Our preliminary results show that it is worth spending more effort to choose a well-designed algorithm and architecture for the motion estimation and the rate control modules before hardware realization. The purpose of this study is not to design a specific MPEG encoding chip but to look into the tradeoffs of various algorithm modules on the entire MPEG chip design. An area-time estimation method is proposed to assist system designers to estimate the time requirement for each functional unit and to find the approximate silicon area of these units. Then, the designer can focus on reducing the chip areas of the dominate modules. Although our tools have limited precision, we still obtain useful guidelines in the MPEG encoder chip design from this study. It is a challenge in the future to develop an optimal way of evaluating the combinations of algorithms, architectures, and layouts at system level.

REFERENCES

[1] S.-C. Cheng and H.-M. Hang, "A Comparison of Block-Matching Algorithms Mapped to Systolic-Array Implementation," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 7, pp. 741-575, Oct. 1997.

[2] S.-C. Cheng and H.-M. Hang, "The Impact of Rate Control Algorithms on Video Codec Hardware Design," in *Proc. Int. Conf. Image Processing*, pp. 807-810, Oct. 1997.

[3] S. Uramoto, Y. Inoue, A. Takabatake, etc., "A 100-MHz 2-D DCT processor," *IEEE J. Selected Areas in Commun.*, vol. 27, no. 4, pp. 492-499, Apr. 1992.

[4] M. T. Sun, T. C. Chen, and A. M. Gottlieb, "VLSI implementation of a 16*16 DCT," *IEEE Trans. Circuits Syst.*, vol 36, no. 4, pp. 610-617, Apr. 1989.

[5] S. M. Lei and M. T. Sun, "An entropy coding system for Digital HDTV applications," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 1, no. 1, pp. 147-154, Mar. 1991.

[6] H. Park and V. K. Prasanna, "Area efficient VLSI architectures for Huffman coding," *IEEE Trans. Circuits Syst.-II: Analog and DSP.*, vol. 40, no. 9, pp. 568-575, Sep. 1993.

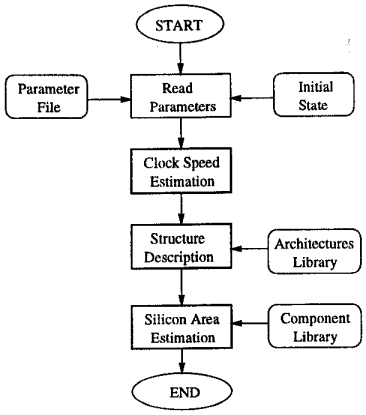

Figure 1: The area-time estimation tool

Table 1: Estimated silicon area for various picture formats

Items	Picture format, search range			
Component	CCIR		CIF	
Gate count (K gate)	47	15	15	7
ME_{full}	487.5	182.4	20.4	9.2
ME_{tss}	24.9	19	10.4	7.7
DCT_1	11	11	9.3	9.3
DCT_2	23	23	19	19
Quan	9	9	7.3	7.3
FMP	9.7	9.7	9.2	9.2
Scan	8.0	8.0	8.0	8.0
VLC_1	3.5	3.5	3.2	3.2
VLC_2	3.7	3.7	3	3
MMU	5.1	4.4	3.1	2.2
$RCTL_{TM5}$	14.7	14.7	8	8
$RCTL_{BP}$	3.9	3.9	1.4	1.4
$RCTL_{OPT}$	294.3	294.3	124.6	124.6

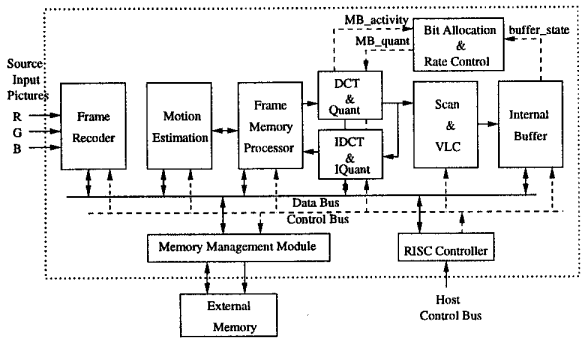

Figure 2: Block diagram of an MPEG encoder chip

Table 2: Internal buffer size of each module in the MPEG encoding chip

Module	Algorithm	Size	Area
ME	FULL	14080	5.6
	TSS	20480	8.2
RCTL	TM5	6370	5.3
	BP	385	0.29
	OPT	57718	54.3
DCT	-	768	0.58
$SCAN$	-	768	0.58
$IDCT$	-	768	0.58

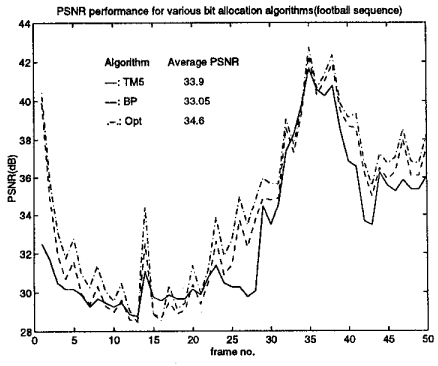

Figure 3: PSNR performance for various rate control algorithms on the *Football* sequence

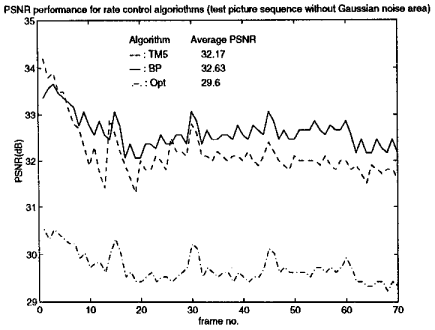

Figure 4: The PSNR of the center of the *Gaussian* sequence