

A COST-EFFECTIVE RECONFIGURABLE ACCELERATOR FOR PLATFORM-BASED SOC DESIGN

Lan-Da Van*, Hsin-Fu Luo, Nien-hsiang Chang, and Chun-Ming Huang

*Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

e-mail: ldvan@cs.nctu.edu.tw

National Chip Implementation Center (CIC), National Applied Research Laboratories, Hsinchu, Taiwan, R.O.C.

E-mail: {brady, nschang, cmhuang}@cic.org.tw

ABSTRACT

In this paper, we propose a cost-effective reconfigurable accelerator for the platform-based system-on-a-chip (SoC) design. Based on the proposed design methodology, the reconfigurable computation array (RCA) can be landed with the features of high usage rate and low hardware cost without sacrificing multimedia computation performance. The RCA consisting of 8 type 1 grouped processing elements (GPE1's), 3 GPE2's and 1 GPE3 is capable of configuring two 16x16-bit multiplication, eight 8x8 multiplication, and sixteen 8-bit absolute operations in different connection topologies. Via the cost-effective RCA, the number of GPEs can be saved up to 25% and the usage rates of the RCA compared with that of [8] for motion estimation (ME), RGB2YUV and DCT/IDCT can be improved by 25%, 18.7%, and 23.9%, respectively.

1. Introduction

With consecutive advances of video algorithm and very-large-scale-integrated (VLSI) technology, an increasing demand of various platform-based SoC designs has evolved in the multimedia field. In the video applications, DSP processor [1-3] or/and accelerator designs [4-5] are the most two off-the-shelf approaches to realize a broad range of video processing algorithms. Because video coding operations such as motion estimation (ME), RGB to YUV (RGB2YUV) conversion, discrete cosine transform (DCT), and finite impulse response (FIR) filter are of computation-intensive loads, we are encouraged to improve the performance of accelerator or DSP processors in no time. On the other hand, due to demands of high-flexibility and mobility for multimedia-communications services, the reconfigurable accelerator and DSP processors [6-9] recently form a new class of architectures. Thus, many significant research efforts focus on developing reconfigurable accelerators or/and DSP processors.

It is well known that TI C'64x [2] is a general-purpose DSP processor, and it is no doubt that the processor is able to achieve very high performance; however, the considerable area issue will affect the cost. In our previous work [3], although the area cost has been improved, this is not compliant with the high-performance demand. Chen *et al.* [8] proposed a high-performance reconfigurable DSP processor to handle higher computation-intensive loads. Nevertheless, due to the poor hardware resource utilization, the usage rate of the computation array is inefficient.

Meanwhile, the data bandwidth and high-interconnection complexity have not pondered thoroughly. Thus, we are motivated to design a cost-effective and low-complexity accelerator for multimedia-oriented SoC design. Based on the four-step design methodology, we can devise a new reconfigurable computation array (RCA) with the features of minimum number of GPE's, high usage rate, and low interconnection complexity without sacrificing performance. Hence, the proposed reconfigurable accelerator is able to meet the high performance demands for various video coding algorithms under the limited hardware resource. The paper is structured as follows. A design methodology of the novel reconfigurable accelerator is exposed in Section 2. In Section 3, configuration benchmark studies are discussed via the proposed RCA. In Section 4, comparison results are tabulated in terms of hardware cost and performance. Most importantly, the hardware cost in terms of the number of GPE's, multipliers, interconnection complexity, and usage rate will be carefully compared among three architectures. The chip implementation of RCA will be debated in the same section. We give a brief conclusion in the last section.

2. Design Methodology of the Novel Reconfigurable Accelerator

Without loss of generality, the simplified platform-based SoC design is deposited in Fig. 1, where the master covers CPU and DSP, and the slave includes accelerator, memory, and IP. The direct memory access (DMA) unit feeds input data stream to the register file (RF) of the accelerator as shown in Fig. 1. Using this approach, we can easily raise above the huge bandwidth requirement. In this paper, devising a cost-effective reconfigurable accelerator is our main goal. The proposed reconfigurable accelerator depicted in the dash line region is composed of the RCA, configuration controller as well as memory, and RF. These functional units of the accelerator are described in more detail as follows.

2.1 Reconfigurable Computation Array (RCA)

The RCA is responsible for the calculation of computation-intensive multimedia algorithms including ME, RGB2YUV, DCT/IDCT, and FIR filter. The RCA design methodology flow chart revealed in Fig. 2 covers four main steps. The steps are explained as follows.

Step 1: Algorithm Exploration

In accordance with a specific class of video algorithms, derive the required computation units that composed of

addition, multiplication, division, or absolute units.

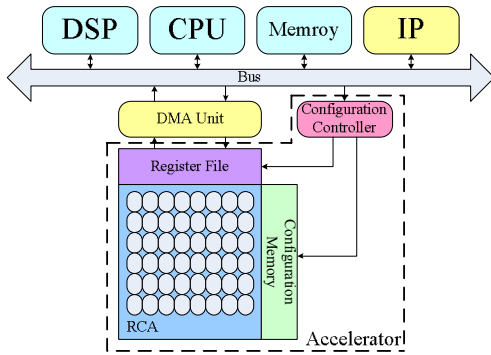


Fig. 1. Simplified platform-based SoC design.

Step 2: GPE Exploration

Explore different VLSI architectures for each computation unit to obtain the one with the highest hardware compatibility among the computation units. The common structures from the computation units are treated as baseline GPE's.

Step 3: RCA Exploration

According to the dominant computation-intensive algorithm, explore different VLSI architectures based on the derived GPE to land the possible RCA structure. In this stage, two control parameters can be used to modify the structure: one is GPE's numbers and the other is connection path. Thus, the primary RCA structure can be determined.

Step 4: Cost & Performance Optimization

Confirm whether the cost meets requirement under satisfying performance constraint. If yes, the final RCA fabric with minimum number of GPE's can be obtained. Otherwise, repeat step 3.

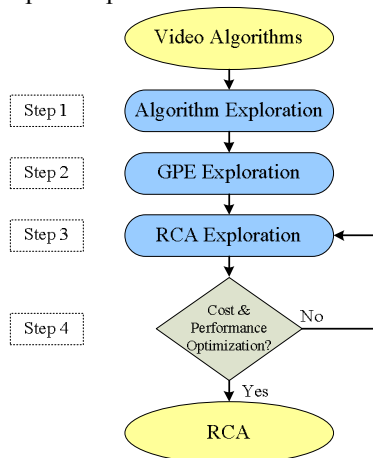


Fig. 2. Flow chart of RCA design methodology.

In the first step, since four algorithms are our object, we can summary that addition, multiplication, and absolute are the main computation unit after the algorithm exploration. Through step 2, it is obviously found that the adder can be used to construct other computation units. From algorithm profiling, ME certainly dominates the computation resource. In the straightforward

design, we at least need 47 GPE's to accomplish one 8x8 block matching in step 3. However, the RCA with 47 GPE's has the disadvantages of high cost and low usage rate in stage 4. Thus, we go back to step 3. Next, we fold the RCA architecture two times and keep the same performance as that of [8]. Finally, we can construct an RCA with 12 GPE's as shown in Fig. 3, where the GPE1 and GPE2 are be easily obtained via grouping 4 PE1's and 4 PE2's of [8] and GPE3 can be regarded as one three 16-bit input adder. Herein, due to the limited pages, we bypass the detailed GPE block diagram. On the other hand, the usage rate of GPE's can be higher than that of [8] via this cost-effective RCA architecture. The RCA only needs to configure GPE's via distinct interconnection to accommodate other computations. In Fig. 3, the interconnection path is responsible for transmitting data from one GPE to another to accomplish the calculation of multimedia-specific algorithms. Since the proposed accelerator focuses on four computation intensive algorithms, the interconnection flexibility will be confined but truly simple compared to that of [8].

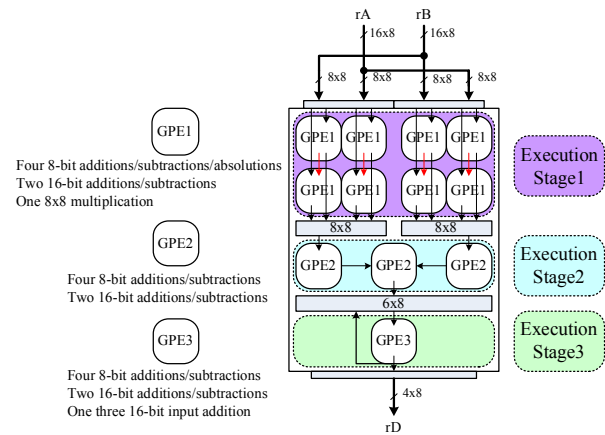


Fig. 3. The proposed reconfigurable computation array.

2.2 Configuration Controller and Memory

The key parameters of the configuration controller are revealed in Fig. 4(a), where rA and rB denote source operand address, and rD represents the destination operand address, and lpc0, lpc1 are loop counter, and bA0, bB0, bD0, bA1, bB1, bD1 denote offset. The function of the configuration controller is to manipulate the selection of the following modes: ME, RGB2YUV, DCT/IDCT, and FIR filter. The pseudo code can be illustrated in Fig. 4(b). The value of each control parameter can be loaded from or stored in the configuration memory as shown in Fig. 1.

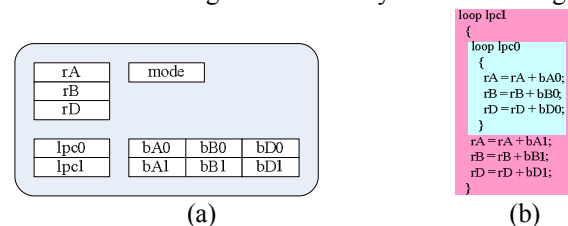


Fig. 4. Configuration controller (a) control parameters, (b) pseudo code.

2.3 Register File (RF)

It is well known that larger RF, few accesses to the main memory can be achieved. Thus, we can tremendously save the operation time. The proposed accelerator requires 2k byte RF with two read ports (16-byte) and one write port (4-byte). The size of RF is dominated by ME.

3. Configuration Studies

In this section, we further explain the configuration models applying our proposed RCA through four benchmarks.

3.1 Motion Estimation (ME)

In [5, 8-9], from the functional analysis for MPEG-4 and H.264/AVC, the most computation-consuming source comes from ME. Via the proposed RCA, we can concurrently calculate sixteen absolute operations between two 8x8 blocks as demonstrated in Fig. 5. The operation procedures are illustrated on the right-hand side of Fig. 5. When we operate one 8x8 block matching, the pointer assignment can be set as listed in Table 1, where lpc0, bA0, bB0, bD0, lpc1, bA1, bB1, and bD1 are 4, 16, 16, 0, 16, -64, 0 and 2, respectively. The detailed operations are described as follows. the loop lpc0 records 4 iterations for one 8x8 block matching. In the lpc1 loop, in order to move rA to the start point of the reference block, we set bA1=-64. rD is used to point the next 8x8 block matching address after adding the value of bD1.

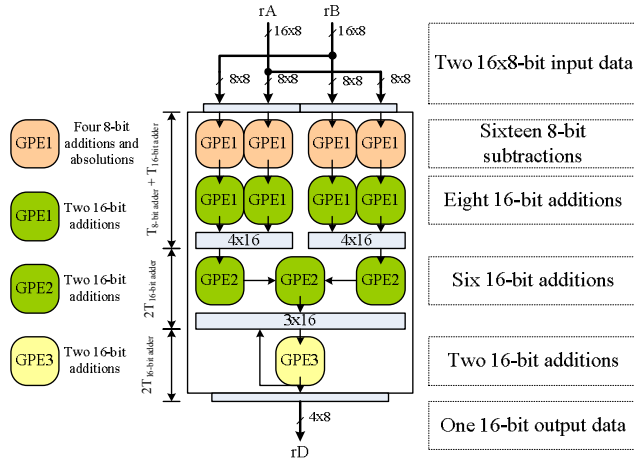


Fig. 5. Configuration of motion estimation.

Table 1: Pointer Assignment of ME

rA = reference image block index							
rB = current image block index							
rD = destination index							
lpc0	4	bA0	16	bB0	16	bD0	0
lpc1	16	bA1	-64	bB1	0	bD1	2

3.2 RGB2YUV with N Pixels

The human eye is more sensitive to luminance than to chrominance. That is why we convert RGB color space to YUV space. Moreover, in most video codecs, the RGB2YUV conversion is one of huge consuming

operations. In similar behavior, the configuration model can be implemented in Fig. 6. The operation notations are appended to the right-hand side at each stage. When one RGB2YUV operation needs to be computed, the values of the pointer of lpc0, bA0, bB0, bD0, lpc1, bA1, bB1, and bD1 can be set 3, 3, 6, 4, 32, 0, -18 and 0, respectively.

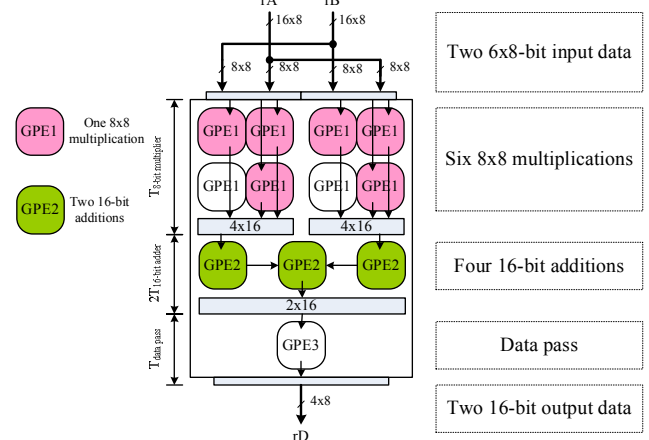


Fig. 6. Configuration of RGB2YUV.

3.3 8x8 DCT/IDCT

The DCT/IDCT is the key computation for MPEG-4 standard. When the 8x8 DCT as shown in Fig. 7 has to be computed, the pointer values of lpc0, bA0, bB0, bD0, lpc1, bA1, bB1, and bD1 can be set 8, 0, 8, 2, 8, 8, 0 and 0, respectively.

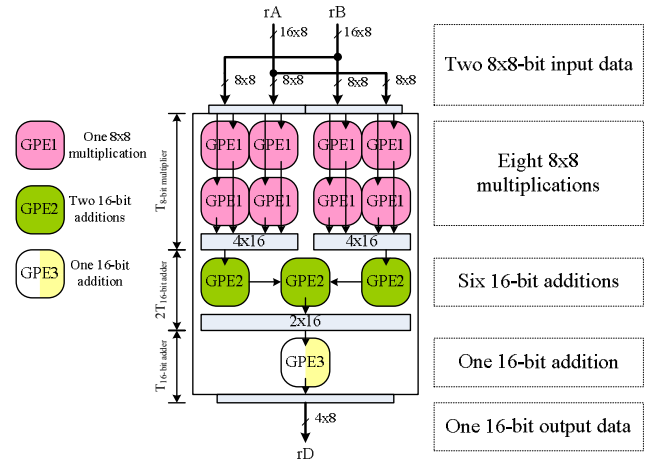


Fig. 7. Configuration of 8x8 DCT/IDCT.

3.4 FIR Filter

The FIR operation has been widely used in H.264/AVC. When the 6-tap FIR filter operation has to be computed, the values of the pointer of lpc0, bA0, bB0, and bD0 can be appointed to 64, 12, 0, and 2, respectively. Since the configuration is very similar to DCT, the detailed procedures can be ignored here.

4. Comparison Results and Implementation

In this section, we give comprehensive comparison results as listed in Table 2 and Fig. 8 in terms of the number of GPE's, multipliers, interconnection complexity,

and performance as well as hardware utilization rate. In terms of the number of GPE's as listed in Table 2, the RCA can be saved by a factor of 25% compared with [8]. In addition, the 12 GPE's can be equivalently regarded as 3 16x16 bit multipliers. Therefore, the proposed RCA owns the lowest hardware cost among three structures [3, 8]. In Table 2, it is manifest that our proposed architecture has superior performance to that of [3] and keeps the same performance as that of [8]. Applying the proposed cost-effective RCA architecture, 25%, 18.7% and 23.9% hardware utilization improvement as shown in Fig. 8 can be landed for ME, RGB2YUV and DCT/IDCT benchmarks, respectively. As a consequence, the higher hardware-utilization efficiency can be achieved. Thus, our proposed reconfigurable accelerator is very cost effective and performance efficient.

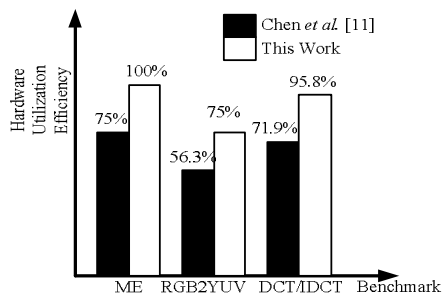


Fig. 8. Hardware utilization efficiency versus benchmarks.

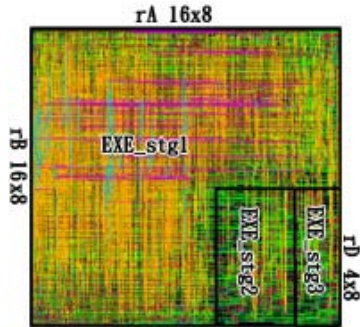


Fig. 9. Chip Layout.

Concerning the chip implementation, it is worth noticing that the RCA mainly dominates the accelerator performance. The active chip layout area of the proposed RCA as shown in Fig. 9 is 320 um x 320 um in TSMC

0.13 um CMOS process. The critical delay time obtained from the static timing analysis (STA) of Synopsys is 12.5 ns (i.e., 80 MHz) under the worst-case condition.

5. Conclusion

We have contributed a cost-effective accelerator based on the novel RCA fabric without sacrificing performance for the platform-based SoC design. Via the proposed RCA, the RCA usage rates for ME, RGB2YUV and DCT/IDCT can be improved by 25%, 18.7%, and 23.9%, respectively, compared with that of [8].

References

- [1] Y. H. Hu, *Programmable Digital Signal Processors: Architectures, Programming, and Applications*, Marcel Dekker Inc., 2002.
- [2] *TMS320C64x DSP Library Programmer's Reference*, Texas Instruments Inc., Apr. 2002.
- [3] L. D. Van, H. F. Luo, C. M. Wu, W. S. Hu, C. M. Huang, W. C. Tsai, "A high-performance area-aware DSP architecture for multimedia codecs," in *Proc. IEEE Int. Conf. Multi. Expo.*, Jun. 2004, vol. 3, pp. 1499-1502, Taipei, Taiwan.
- [4] V. A. Chouliaras, J. L. Nunez, D. J. Mulvaney, F. S. Rovati, D. Alfonso, "A multi-standard video accelerator based on a vector architecture," *IEEE Trans. Consumer Electronics*, vol. 51, pp. 160-167, Feb. 2005.
- [5] S. Y. Chien, Y. W. Huang, C. Y. Chen, H. H. Chen, L. G. Chen, "Hardware architecture design of video compression for multimedia communication systems," *IEEE Communications Magazine*, pp. 123-131, Aug. 2005.
- [6] R. Hartenstein, "Trends in reconfigurable logic and reconfigurable computing," in *Proc. IEEE Int. Conf. Electronics, Circuits, and Systems*, 2002, pp. 801-808.
- [7] H. Zhang, V. Prabhu, V. George, M. Wan, M. Benes, A. Abnous, J. M. Rabaey, "A 1-V heterogeneous reconfigurable DSP IC for wireless baseband digital signal processing," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 1697-1704, Nov. 2000.
- [8] L. H. Chen, Oscar T. C. Chen, R. L. Ma, "A high-efficiency reconfigurable digital signal processor for multimedia computing," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2003, vol. 2, pp. 768-771, Bangkok, Thailand.
- [9] L. H. Chen, Oscar T. C. Chen, T. Y. Wang, C. L. Wang, "An adaptive DSP processor for high-efficiency computing MPEG-4 video encoder," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2004, vol. 2, pp. 157-160, Vancouver, Canada.

Table 2: Cost and Performance Comparison Results

Feature		Van et al.	Chen et al.	This Work
Reconfigurability		No	Yes	Yes
Hardware Cost	# of GPE's	0	16 (Equivalent)	12
	# of Multipliers	4 16x16-bit	0	0
	Interconnection Complexity	Low	High	Low
Performance	Motion Estimation (Pixel per Execution Stage)	8	16	16
	RGB2YUV with N Pixels (Execution Cycles)	$3N/2$	$3N/2$	$3N/2$
	2-D DCT/IDCT for $N \times N$ pixels (Execution Cycles)	$N^3/2$	$N^3/4$	$N^3/4$