RECONFIGURABLE AND PROGRAMMABLE MINIMUM DISTANCE SEARCH ENGINE FOR PORTABLE VIDEO COMPRESSION SYSTEMS

Chang-Ki Kwon

Memory R&D Division, Hyundai Electronics Co.,Ltd.
E-mail: ckkwon@sr.hei.co.kr

Kwyro Lee, Senior Member of IEEE

Department of EECS, KAIST,
E-mail: krlee@ee.kaist.ac.kr

ABSTRACT

A Reconfigurable and Programmable Minimum Distance Search Engine (RP-MDSE) is presented as a hardware accelerator for portable video compression systems. Our architectural-level estimation shows that a reconfigurable structure with memory-merged logic and low-power processing elements can achieve high throughput as well as dramatic power savings. Based on the measurement results of our prototype chip fabricated with a 0.6μm single-poly triple-metal CMOS technology, the estimated one dissipates 32~66% less power than the other chips for H.26x motion estimation. Moreover, it operates as fast as the recent vector quantization chip, while having a small code-book of 16 dimensional 8 code-vectors.

1. INTRODUCTION

Recently the market for wireless multimedia communication systems and portable video terminals has been on the rise. Thus, it has become of great importance to implement low-power portable video compression systems.

Among all the video compressing process, "nearest neighbor search" such as Vector Quantization (VQ) [1] and Motion Estimation (ME) [2] requires huge amount of computation and large portion of power. This leads us to develop a low power hardware accelerator, which finds the vector nearest to an input one among a large number of reference vectors after efficiently adapting itself to the respective application.

In this paper, we propose a novel reconfigurable structure tuned to the nearest neighbor search application. We also present our VLSI implementation by using Memory-Merged Logic (MML) [3] for further improvement of performance such as power and area.

2. ARCHITECTURAL ESTIMATION

In this section, our architectural-level power estimation for Full Search ME (FS-ME) systems and Full Search VQ (FS-VQ) systems are described.

Firstly, we estimate architectural-level power consumption of the FS-ME systems. The system is organized with three basic components: a previous frame memory, a current frame memory and a motion estimator. The architecture can be classified into two generic types shown in Fig. 1.

Wuytack [4] presented his power exploration methodology based on the observation that for data-dominated video applications the power consumption is dominated by the memory accesses. Thus, he described how to optimize memory architecture with a simple data-path processor, as shown in Fig. 1(a). However, other researchers [5][6] have focused on how systolic array-processors can be utilized to increase computational power together with reducing the overlapped data transfers.

Now let us compare quantitatively the power consumption of the two types shown in Fig. 1, given the following specifications: the FS-ME with gray-scaled pseudo-random CIF images (HxV) of 352x288, a reference block size (NxN) of 8x8 pels, a frame rate (fr) of 30 Hz and a maximum displacement of p.

For architectural-level power estimation in practical cases, we assume that the supply voltage is fixed. In addition, it is assumed that the clock frequency (fclk) is scaled linearly below the maximum frequency (fmax) at which external SRAM is accessed, and that fclk can be gated according to the power-down strategies. Then, fmax is only needed to compute the minimal number of parallel chip-sets.

For the SRAM, we have used the power model of Fujitsu's 1Mbit-SRAM [7] whose maximum frequency is 100MHz. The other well-developed power models of embedded RAM [8], registers [8], and data-paths [4], have been also adopted. All the models have been modified for a 0.5-μm CMOS process, operating at 3.3V, by Constant E-Field scaling [9].

Fig. 2 shows the detailed power dissipation of Fig.1(a)~(b), neglecting the dashed boxes due to their negligible contribution. We have observed that most of power is dominated by internal memory access in Fig. 2(a). This is because the off-chip memory access exceeds the maximum frequency as the searching range increases, as pointed out in [4]. In Fig. 2(b), notice that the systolic structure tuned to the FS-ME saves much power by an order of magnitude than the general-purpose processor-based system, even though it can't avoid the overlapped data-transfer of search areas between adjacent reference blocks.

Similarly, for the FS-VQ systems, we observed that for dramatic power savings our RP-MDSE must involve many distributed arithmetic elements closely connected to internal memories in parallel on a chip, as previously presented in [3].

Consequently, we have noticed the following facts. Firstly, our MDSE must have reconfigurable programmable architecture tuned to the respective application such as FS-ME and FS-VQ. Secondly, all the Processing Elements (PE's) should be optimized in circuit level for further improvement of performance such as power dissipation and area.

0-7803-6685-9/01/$10.00©2001 IEEE

3. VLSI IMPLEMENTATION

In this section, overall architecture of our RP-MDSE and our circuits are described in detail.

3.1 Reconfigurable and Programmable Structure

Fig. 3(a) shows the overall structure of our RP-MDSE, which can be used as three configurations such as a conventional SRAM, a MDSE for FS-VQ, and a MDSE for FS-ME.

Fig. 3(b) shows the main reconfigurable modes. According to the respective application, if external command signals are inserted into the global control logic, our RP-MDSE is then properly reconfigured with internal control signals such as reset, cfg1, and cfg2.

Firstly, the functional blocks for a SRAM are the following: address decoders, I/O interface circuits, sense amplifiers, SRAM cells in PE's, and Shift Latch Arrays (SLA's). All the blocks operate like a conventional SRAM.

Secondly, our RP-MDSE for FS-VQ consists of MML in the PE's and Digital/Analog-Mixed Winner-Take-All Circuits (DAM-WTAC) [3]. After the distances between a component of a N-dimensional input vector and pre-stored components of N-dimensional code-vectors are calculated in parallel at one column, the values are stored in parallel row by row into the parallel accumulators, respectively. Then, these operations are repeated sequentially in the column-wise order to the final column. Eventually, our DAM-WTAC select the position of the minimum distance code-vector.

Finally, our RP-MDSE for FS-ME is organized with cross-striped P/N-PE's, SLA's, a serial Adder Chain (AC), and a Motion Vector Calculator (MVC). The detailed operations are described as follows.

Fig. 4(a) describes the simplified systolic structure with the reference block data x(i,j) and the search area data y(i+m,j+n). The reference data are already stored and remain fixed inside the PE's, respectively. Then, the search area data are serially inserted into the bottom right P-PE. They are also shifted in the systolic arrays and SLA's along the narrow solid line (see Fig. 3(a)). Consequently, by the programmable SLA's, the search area data can be effectively fed, as shown in Fig. 4(a). Notice that the input sequences of the search area data, shifted by only one half-clock row by row, are fed into the systolic arrays by the clock. Moreover, notice that our MML per PE that performs both of the Absolute Difference Calculation (ADC) and the Summation (SUM) in one clock cycle, as shown in Fig. 4(b). Eventually, all the PE's perform the ADC's in parallel by the column and send the partial sums of the values in sequence by the column to the serial adder chain, to compute the Mean Absolute Difference (MAD). The MAD's are fed serially into the MVC to find the motion vector. Thus, for finding one motion vector per reference block, the total processing clocks of approximately $(N+2p)^2$ are required.

Notice that our systolic architecture is similar to that of [5][6] except our power/area-efficient PE circuits.

3.2 P/N-type Processing Elements

Our P/N-PE consists of the modified SSTC p-latches [10] and programmable MML circuits [3]. The detailed circuit of a 1-bit P-PE is shown in Fig. 5. The b-bit P-PE can be constructed by cascading b 1-bit P-PE's. For all the N-PE's, the underlined rectangles illustrated in Fig. 5 must be replaced by the underlined rounded-rectangles. Thus, for FS-ME, all the N-PE's do the same operation of the P-PE's at the opposite clock phase.

4. EXPERIMENTAL RESULTS

With a 0.6 μm single-poly triple-metal CMOS process, we have implemented a prototype RP-MDSE chip for FS-VQ as well as FS-ME. It involves 6 bit – 16 word PE's, programmable SLA's with maximum displacement of 2, 6 bit – 4 word DAM-WTAC, I/O circuits, and a simple global control logic (within the limited available area).

The measured performance of the chip is summarized in Table I, and the microphotograph is shown in Fig. 6.

Fig. 7 compares the measured performance of our prototype chip with the existing FS-ME chip data [11][12] and the recent FS-VQ chip data [13][14]. Most of the chips were implemented with a standard 0.5 ~ 0.9 μm CMOS process, operating at 3 ~ 5V supply. Based on the measurement results of our prototype chip, the estimated one for H.26x FS-ME dissipates 32 ~ 66% less power than the Lin's [12] at the expense of about 3 ~ 5 times larger area. Notice that this area penalty could be considerably reduced with careful layout since our RP-MDSE reduces the number of transistors by using our power/area-efficient MML. Moreover, for FS-VQ, it operates as fast as the Nakada's [14], while having a quite small code-book of 16 dimensional 8 code-vectors.

5. CONCLUSION

A low-power RP-MDSE has been presented as a hardware accelerator for portable video compression systems.

Our architectural-level power estimation results show that a reconfigurable system with memory-merged logic and low-power processing elements can achieve high throughput as well as dramatic power savings, compared with the general-purpose processor-based systems. Based on the measurement results of the fabricated prototype chip, the estimated one dissipates 32 ~ 66% less power than the other H.26x FS-ME chip. However, it seems to be less useful than the recent VQ chip due to the smaller code-book size of 16 dimensional 8 code-vectors.

It is expected that the design techniques as well as the chip will be useful for low-power multimedia search application such as speech recognition, pattern matching, multimedia database search, and so forth.

ACKNOWLEDGEMENT

This work was partly supported by MICROS Research Center and IDEC.

REFERENCES

[1] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Boston, Kluwer Academic Publishers, 1992.

[2] T. Chen, "Recent Development in Video Coding Standards for Multimedia Communications: H.261, H.263, and H.263+", in *ISCAS' 97 Tutorial on MPEG-4*, June 1997.

[3] C. K. Kwon and Kwyro Lee, "A Low Power Minimum Distance 1D-Search Engine Using Hybrid Digital/Analog Circuit Techniques", *ISCAS*, pp.214-217, June 1999.

[4] S. Wuytack, et als., "Power Exploration for Data Dominated Video Applications", *IWLPED*, pp.359-364, Aug. 1996.

[5] T. Komarek and P. Pirsch, "Array Architectures for Block Matching Algorithm" *IEEE T.CAS*,pp.1301-1308,Oct. 1989.

[6] C. H. Hsieh and T. P. Lin, "VLSI Architectures for Block Matching Motion Estimation Algorithm", *IEEE T. on CAS for Video Tech.*, pp. 169-175, June 1992.

[7] T. Seki, et al., "A 6ns 1Mb CMOS SRAM with Latched Sense Amplifier", *IEEE J. of SSC*, pp. 478-483, April 1993.

[8] D. Lidsky and J. M. Rabaey, "Early Power Exploration – A World Wide Web Application", *DAC*, pp. 27-32, June 1996.

[9] R. H. Dennard, et al., "Design of Ion Implanted MOSFETs with Very Small Physical Dimensions", *IEEE J. of SSC*, pp. 256-266, Oct. 1974.

[10] J. Yuan and C. Svensson, "New Single-Clock CMOS Latches and Flipflops with Improved Speed and Power Savings", *IEEE J. of SSC*, pp.62-69, Jan. 1997.

[11] G. K. Yeh, et al., "A Low Power Video Motion Estimation Array Processor", *SOVC*, pp. 162-163, June 1996.

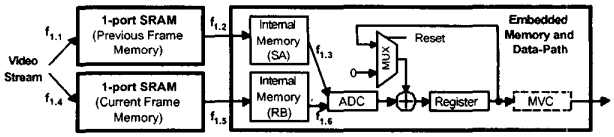
[12] H. D. Lin, et al., "A 14-Gops Programmable Motion Estimator for H.26x Video Coding", *IEEE J. of SSC*, pp. 1742-1750, Nov. 1996.

[13] K. Kobayachi, et al., "A Memory-based Parallel Processor for Vector Quantization" *ESSCIRC*, pp.184-187, Sept. 1996.

[14] A. Nakada, et al., "A Fully Parallel Vector Quantization Processor for Real-Time Motion Picture Compression", *IEEE J. of SSC*, pp. 822 -830, June 1999.

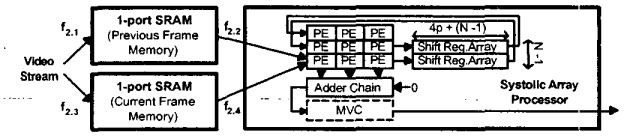
Table I. The measured performance of the prototype 6bit-16 word RP-MDSE chip.

	6 bit – 4 word MML	6 bit – 4 word DAM-WTAC	6 bit – 16 word PE's
Technology	0.6 μm 1-Poly 3-Metal CMOS	2.0 μm (minimum length)	0.6 μm 1-Poly 3-Metal CMOS
Supply Voltage	3 ~ 5 V	3 ~ 5 V	3 ~ 5 V
Delay Time	< 12 ns (simulated)	< 60 ns @3.3V	< 16 ns @3.3V
Power Dissipation	4.5 mW @3.3V/10MHz	(static) 0.3 mW @3.3V	19.2 mW @ 3.3V/10MHz
Total Area	-	0.13x0.35(mm^2)	3.5x1.2(mm^2)



(a) Type A – MDSE

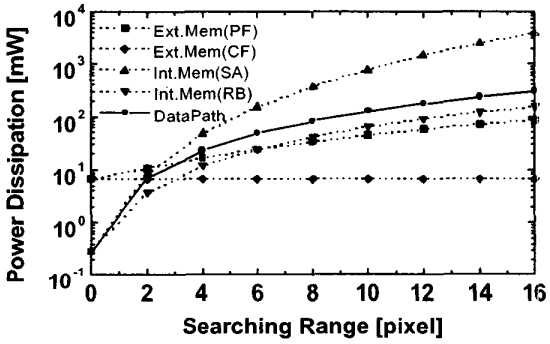
($f_{1.1}$=(H/N)x(V/N)xN^2xfr, $f_{1.2}$=(H/N)x(V/N)x$(N+2p)^2$xfr, $f_{1.3}$=(H/N)x(V/N)xN^2x$(2p+1)^2$xfr, $f_{1.4}$=(H/N)x(V/N)xN^2xfr, $f_{1.5}$=(H/N)x(V/N)xN^2xfr, $f_{1.6}$=(H/N)x(V/N)xN^2x$(2p+1)^2$xfr)



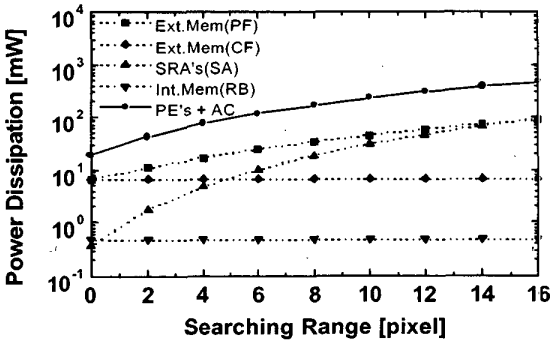
(b) Type B – MDSE

($f_{2.1}$=(H/N)x(V/N)xN^2xfr, $f_{2.2}$=(H/N)x(V/N)x$(N+2p)^2$xfr, $f_{2.3}$=(H/N)x(V/N)xN^2xfr, $f_{2.4}$=(H/N)x(V/N)xN^2xfr)

Figure 1. Block diagrams of Minimum Distance Search Engine (MDSE) for the exhaustive Full Search Motion Estimation (FS-ME).

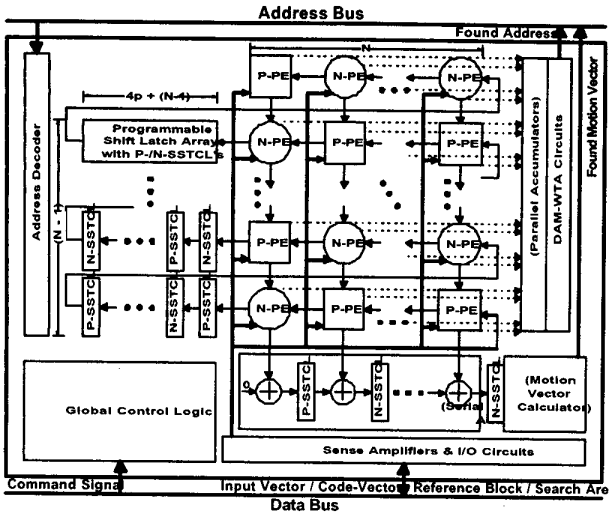


(a) A breakdown of average power of Type A - MDSE

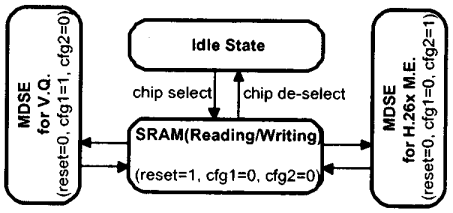


(b) A breakdown of average power of Type B - MDSE

Figure 2. The detailed power dissipation of Fig. 1(a)~(b).

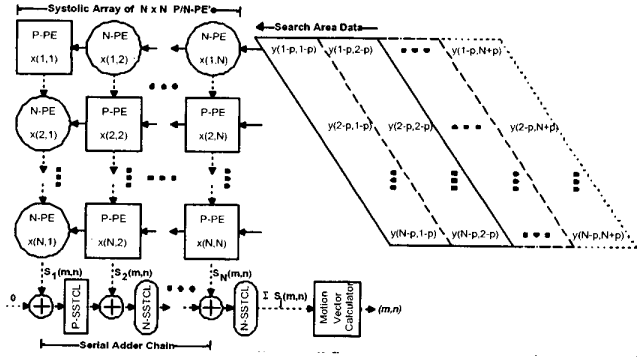


(a) The overall architecture of our RP-MDSE



(b) The main reconfigurable modes of our RP-MDSE

Figure 3. Our Reconfigurable and Programmable MDSE (RP-MDSE)



(a) The simplified structure of our RP-MDSE for FS-ME.

P-PE :	ADC	SUM	ADC	SUM	ADC
N-PE :	SUM	ADC	SUM	ADC	SUM

(b) Timing diagram and operation of P/N-PE's for FS-ME.

Figure 4. The simplified operation per PE of our RP-MDSE for FS-ME.

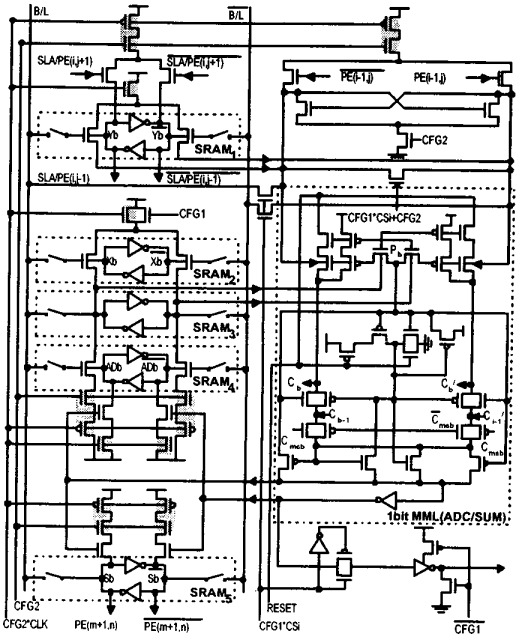


Figure 5. The proposed 1-bit P-PE circuit.

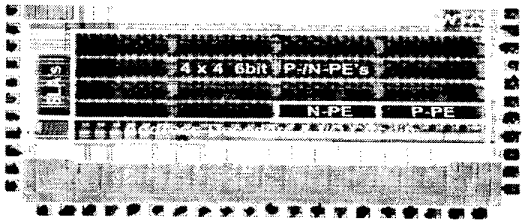


Figure 6. The microphotograph of the fabricated prototype RP-MDSE chip.

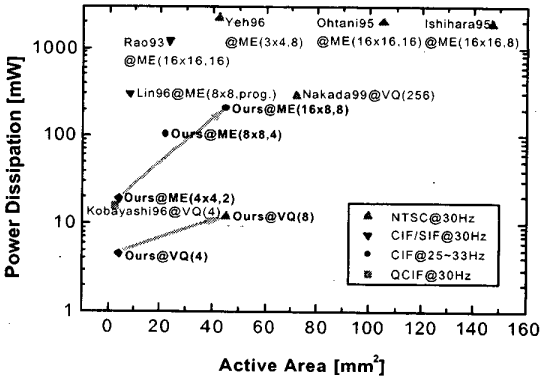


Figure 7. Performance of our chips, compared with other implementations. Notice that ME(*mxn*, *p*) means the FS-ME with a reference block size of *mxn* and a maximum displacement of *p*, and that VQ(*k*) implies the FS-VQ with a code-book of 16 dimensional *k* code-vectors.