A Novel Architecture of Local Memory for Programmable SIMD Vision Chip

Zhe Chen, Jie Yang, Cong Shi, Nanjian Wu*

State Key Laboratory for Superlattices and Microstructures, Institute of Semiconductors,
Chinese Academy of Sciences, Beijing, 100083, China
* Email: nanjian@red.semi.ac.cn

Abstract

This paper presents a novel architecture of the local memory for the programmable SIMD vision chip. The memory architecture consists of 8 × 8 local memory cells, among which each 8 static latches in the master stage share one dynamic latch in the slave stage. The local memory performs single bit read and write in each clock cycle, and the compact area of 14.33 μm^2/bit increases the integration level of the processor. A prototype chip with 64 × 64 processing units has been manufactured in 0.18 μm CMOS technology. Five types of local memory architecture have been designed, and an 8-bit input data buffer based on dedicated latch structures has been designed as the input data buffer for each processing unit. Test results show that the presented structure is suitable for real-time computer vision applications such as the edge detection at the speed of 1000 fps.

Index Terms — image processing, latches, memory architecture, SIMD, vision chip.

1. Introduction

Vision chips are devices integrating image sensors and processing circuits. Compared to the conventional vision system with separate sensing and processing components, vision chips benefit from high performance computing without I/O bottleneck as well as small silicon area and reduced power consumption [1], [2], so these devices show good prospects in high speed image processing applications such as industrial machine vision, target tracking, and real-time surveillance. Programmable single instruction multiple data (SIMD) processors designed for vision chips show high speed performance in mid-level and low-level image processing [3]. In further, vision chips with multi-level architecture including MPU gain enhanced capability in high-level image processing algorithms, such as the PPED pattern extraction [4] and the fast traffic lane detection [5].

This paper presents a novel architecture of the local memory for the programmable SIMD vision chip. In this architecture, the local memory combines static and dynamic latch structures within a compact area, and performs single bit read and write operations as the random access flip-flop register. The basic memory cell occupies 14.33 μm^2 and the 64-bit internal memory occupies 916.9 μm^2, which shows benefits in improving the integration level of the SIMD processor. A dynamic transmission gate edge-triggered input data buffer is also presented to fulfill the task-level parallelism of the global data flow and the massively parallel computation.

This paper proceeds as follows. In section 2, the architecture of the chip is introduced. In section 3, the circuits design and simulation results are presented. In section 4, the chip implementation and performance are shown. Finally, we draw the conclusion in section 5.

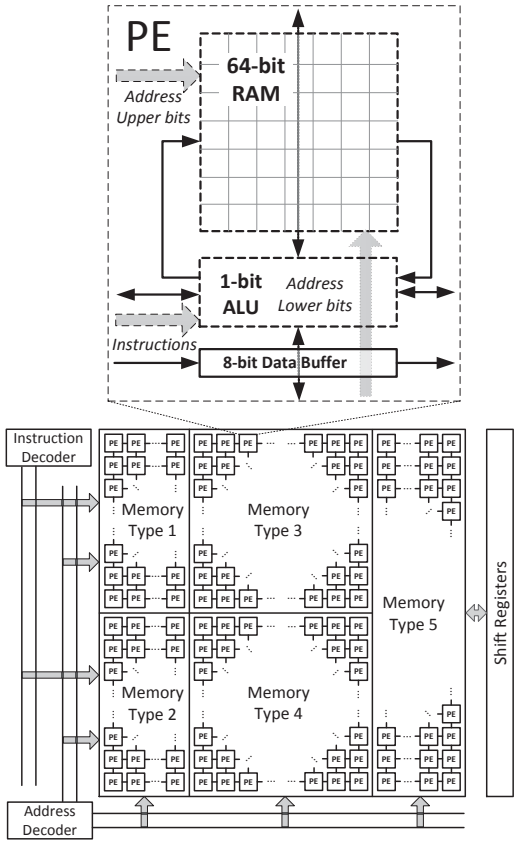

Figure 1. Chip architecture

978-1-4673-6417-1/13/$31.00 ©2013 IEEE

2. Architecture

Figure 1 shows the architecture of the programmable SIMD processor. It consists of 64 × 64 processing elements (PEs), arranged in a 4-connected 2-dimension (2D) array. Each PE consists of a 64-bit local memory and a 1-bit arithmetic-logic unit (ALU). The local memory is arranged as an 8 × 8 2D array with 64 basic memory cells. The ALU performs a 1-bit operation from ADD, INV, AND, and OR in each clock cycle. The PE array is operated under global control signals including arithmetic-logic instructions and memory addresses. Instructions are decoded into two pairs of operand selection signals and one logic operation selection signal. Memory addresses are decoded into upper and lower 8-bit address selection signals. Address selection signals are directed across the local memory in perpendicular directions so that routing resources are saved. The PE array is divided into five sub arrays according to their different local memory types. Array 1 and 2 utilize the single-stage latch architecture as the local memory. Array 3 and 4 add shared dynamic latches in the slave stage based on the former designs. At last, array 5 is designed with the most conservative two-stage register architecture for comparison.

3. Circuits design

3.1 Static latch and dynamic latch

Figure 2 shows two basic memory cells, the 7-T static latch [4] and the 5-T dynamic latch. The 7-T static latch utilizes a feedback PMOS transistor to keep the stored signal level. The advantage of this latch is its stable storing capability and compact size. Based on that, we proposed a 5-T dynamic latch in Figure 2(b). The dynamic latch utilizes a weak PMOS transistor to recover the signal level and reduce the leakage current.

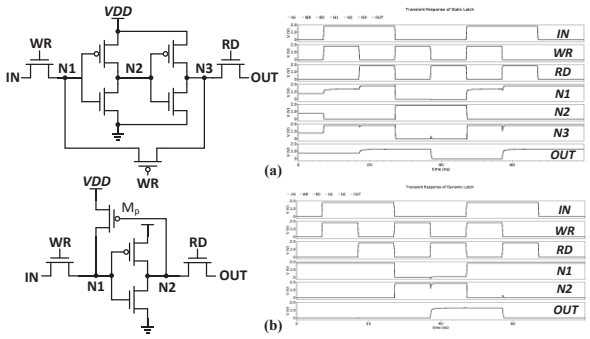

Figure 2. (a) Static latch cell and simulation waveform. (b) Dynamic latch cell and simulation waveform.

We simulate these two basic memory cells with a 20 fF capacitance load. In the transient simulation of the static latch, N3 slightly drops down when RD changes from 0 to 1. This is caused by the charge sharing between OUT and N3. However, N2 keeps 0 when N3 drops down, so it drives N3 back to the original high signal level in very short time. In the transient simulation of the dynamic latch, the transient waveform shows that the transmission transistor does not cause a threshold voltage drop at N3. This is because the PMOS transistor M_p recovers the signal level to VDD soon after N3 turns from low to high. It is noticed that the output signal of the dynamic latch is inverted because the dynamic latch is made up of only one inverter inside.

3.2 Master-slave local memory architecture

In order to realize the read and write operations for a single bit simultaneously, we propose the master-slave memory architecture based on the static and dynamic latch, shown in Figure 3. The master stage consists of 8 static latches, while the slave stage is one dynamic latch shared by 8 bit cells to save area. The feedback PMOS transistor in the master stage helps store the signal state when the WR signal is not activated and the clock level is low. At the positive edge of the clock, one reading path is selected towards the slave stage so that its state is refreshed. Since the read address is synchronized with the clock, the dynamic latch refreshes its state in every clock period, so its state is well reserved.

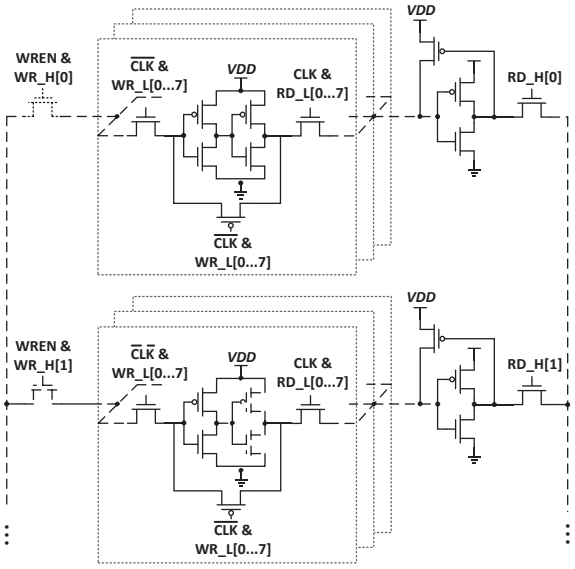

Figure 3. Schematic of the static-latch-master dynamic-latch-slave two-stage register

In a similar way, we designed the dynamic-latch-master dynamic-latch-slave two-stage register. It should be emphasized that this architecture need the feedback PMOS transistor in the master dynamic latch be weak enough to allow input data 0 to be written in.

In order to demonstrate the advantage of the proposed architecture, we also came up with a static-latch-master static-latch-slave two-stage register, as shown in Figure 4. The added inverter eliminates the charge sharing inside the static latch.

Five types of local memories are summarized in Table 1. The area of the static-latch-master dynamic-latch-slave register is 25% smaller than the static-latch-master static-latch-slave two-stage register.

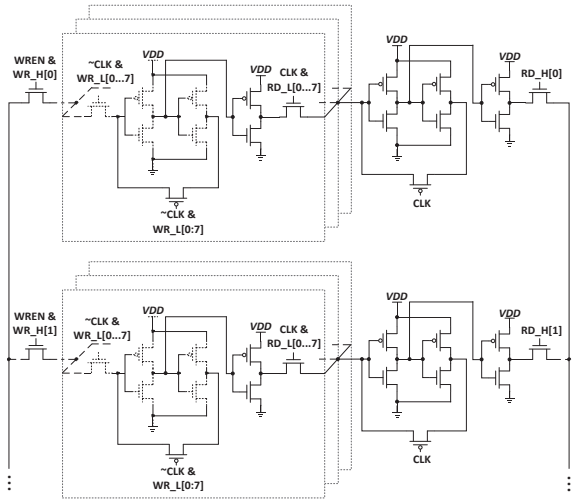

Figure 4. Schematic of the static-latch-master static-latch-slave two-stage register

Table 1. Summary of local memory types

Local Memory Type	PE array	Area / 64-bit
static latch	32 × 16	897.5 μm^2
dynamic latch	32 × 16	783.5 μm^2
static-dynamic register	32 × 32	916.9 μm^2
dynamic-dynamic register	32 × 32	783.9 μm^2
static-static register	64 × 16	1229.9 μm^2

3.3 Data buffer

The PE circuit consists of different memory types and registers, respectively, as shown in Figure 5. We added an 8-bit input data buffer to each PE in order to perform the global data flow and the massively parallel operation simultaneously. The buffer is realized by a full custom dynamic transmission gate edge-triggered shift register chain to reduce the circuit area. The shift register refreshes its stored signal level in each clock cycle when a new image data flows into the PE array, so the signal level is well reserved.

4. Implementation

The proof-of-concept chip has been fabricated in a standard 0.18μm 1P6M CMOS technology. Figure 6 shows the microphotograph of the PE sub array 3 with the proposed static-latch-master dynamic-latch-slave two-stage architecture in this paper, and a complete PE layout in this array. In the layout, basic structures including the ALU, the local memory, and the input data buffer are marked separately. With the use of the layout multiplexing technique, the proposed static-latch-master dynamic-latch-slave two-stage register cell reaches the compact area of 14.33μm^2/bit, and the 64-bit local memory based on the proposed architecture occupies 916.9μm^2, or 41.4% of the complete PE area.

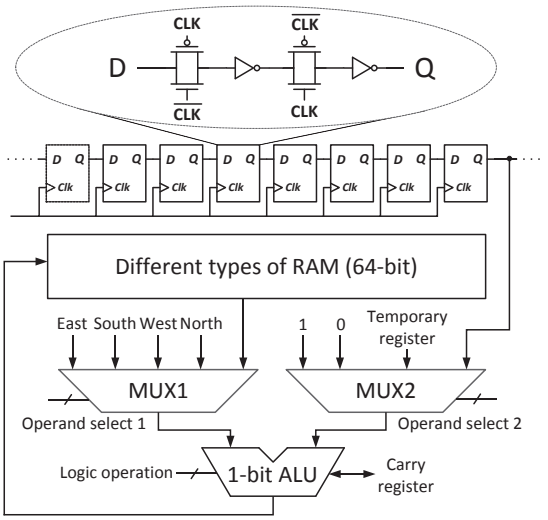

Figure 5. Schematic of PE with different local memory types and registers.

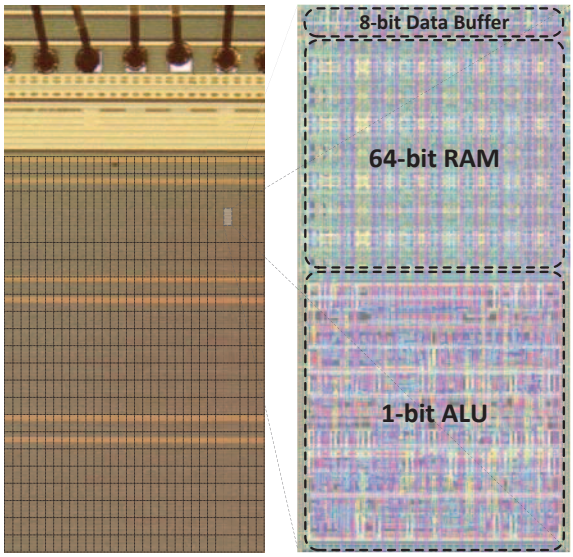

Figure 6 Microphotograph of the type-3 PE sub array and the layout of a complete PE

We set up the test platform for the prototype chip with an external FPGA board. The FPGA generates instructions for the SIMD processing operations.

Table 2. Performance comparison with a synthesized design and other vision chips

Reference	This work	Synthesized design	ASPA [1]	SPE [6]
Technology, μm	0.18	0.18	0.35	0.35
PE array	64 × 64	64 × 64	19 × 22	64 × 64
PE cell size, μm^2	66.3 × 33.4	~3400	100 × 117	67.4 × 67.4
Transistor number/PE	864	1224	460	N/A
PE local memory	72-bit register	64-bit SRAM	64-bit DRAM	24-bit register
Memory cell size, μm^2	14.33	41.1	3.6 × 8.7	N/A

Under this test platform, we test the prototype chip with basic image processing algorithms, such as the image edge detection algorithm. The process flow is described as follows. First, an original 8-bit grayscale image was resampled to the resolution of 64 × 64. Secondly, the resized image was loaded to the PE array row by row under the FPGA control. Thirdly, selected PE sub arrays, namely sub array 3 and 5 in our test, realized the edge detection algorithm by operating in a SIMD manner under the global instructions. Finally, the processed image shifted from the PE array to the I/O interface and was transmitted to the test platform. The edge detection speed reaches 1000 fps and the test results are shown in Figure 7.

The main advantage of the presented architecture is the compact area of the local memory. The full custom design of the local memory is compared to the peer design with the same chip architecture synthesized by the RTL compiler and the memory compiler EDA tools. In the synthesized design, the local memory is realized by dual-port SRAM with the 0.18 um CMOS technology. The area of the local memory and PE circuits are estimated according to the synthesis results. Table 2 shows the comparison of the performance with the synthesis results and other reported vision chips.

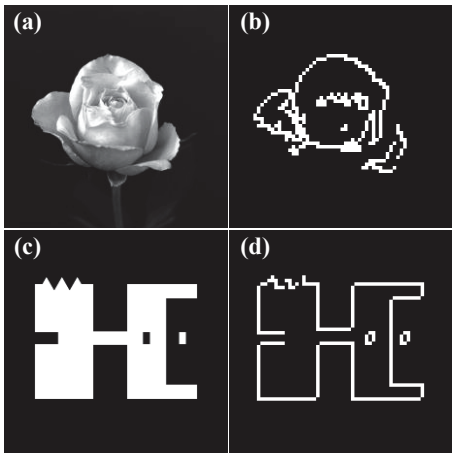

Figure 7 Experiment results with the prototype chip. (a) Original image of a rose. (b) Results of edge detection of (a). (c) Original image of a component. (d) Results of edge detection of (c).

5. Conclusion

In this paper, a programmable SIMD processor based on 64 × 64 PE array designed for vision chip was proposed. The PE was implemented with novel local memory architecture. The static-latch-master dynamic-latch-slave two-stage register works in the same way as a flip-flop register and its area for single bit storage is 14.33 μm^2, which is 25% smaller than the static-latch-master static-latch-slave two-stage register. The prototype chip has been fabricated in a 0.18 μm CMOS technology. Test results show that edge detection algorithm is fulfilled within 1ms, which makes an important sense for practical applications such as the industrial machine vision.

References

[1] A. Lopich and P. Dudek, "A SIMD cellular processor array vision chip with asynchronous processing capabilities," *IEEE Trans. Circuits Syst. I, Reg. Papers,* vol. 58, no. 10, pp. 2420-2431, Oct. 2011.

[2] W. Miao, Q. Y. Lin, W. C. Zhang, and N. J. Wu, "A programmable SIMD vision chip for real-time vision applications," *IEEE J. Solid-State Circuits,* vol. 43, no. 6, pp. 1470-1479, Jun. 2008.

[3] Q. Y. Lin, W. Miao, W. C. Zhang, Q. Y. Fu, and N. J. Wu, "A 1,000 frames/s programmable vision chip with variable resolution and row-pixel-mixed parallel image processors," *Sensors,* vol. 9, no. 8, pp. 5933-5951, Aug. 2009.

[4] W. C. Zhang, Q. Y. Fu, and N. J. Wu, "A programmable vision chip based on multiple levels of parallel processors," *IEEE J. Solid-State Circuits,* vol. 46, no. 9, pp. 2132-2147, Sept. 2011.

[5] Y. J. Li, W. C. Zhang, and N. J. WU, "A novel architecture of vision chip for fast traffic lane detection and FPGA implementation," in *IEEE 8th Int. Conf. on ASIC (ASICON '09)*, 2009, pp. 917-920.

[6] T. Komuro, S. Kagami, and M. Ishikawa, "A dynamically reconfigurable SIMD processor for a vision chip," *IEEE J. Solid-State Circuits,* vol. 39, no. 1, pp. 265-268, Jan. 2004.