An 80×80 general-purpose digital vision chip in 0.18 μm CMOS technology

Alexey Lopich and *Piotr Dudek*
School of Electrical and Electronic Engineering
The University of Manchester
United Kingdom
{a.lopich, p.dudek}@manchester.ac.uk

Abstract— **In this paper we present an implementation of the asynchronous/synchronous processor array (ASPA2) – a digital SIMD vision chip. The chip has been fabricated in a 0.18 μm CMOS process and comprises 80×80 array of pixel processors. The architecture of the chip is overviewed, the design of the processing cell is presented and implementation issues are discussed. At 75 MHz ASPA2 demonstrates 373 GOPS/W power and 871 MOPS/mm^2 area efficiency, making it suitable for the design of high speed and low power vision systems.**

I. Introduction

The ongoing research and development of smart image sensors indicates that they can facilitate computationally demanding image pre-processing at speeds suitable for real time applications. The idea of combining image sensor and processor on a single chip has been explored for more than a decade [1]. Successful implementations of pixel-parallel processor arrays, where every cell combines sensing and processing element, have been reported in [2-4]. The majority of presented vision chips operate in synchronous mode according to SIMD paradigm. However, it has been shown that in a number of pre-processing applications synchronous approach is not the most optimal [5]. A number of architectures based on asynchronous or mixed processing has been presented [6, 7] and proven to be efficient at implementing complex image processing routines.

Another important aspect, which affects overall design and performance of the device, is the signal processing paradigm upon which the processing cell operates (analogue, digital or mixed-signal). Although the analogue approach has been shown to be more power and area efficient [2, 4, 8] compared with its digital counterparts implemented in the same technology node, it should be noted that the referenced designs have been implemented in relatively old technologies and aforementioned efficiency is not preserved with continuing scaling of fabrication process. Robustness and stronger noise immunity make the digital approach more favourable in nanoscale CMOS technologies.

In our previous work [8] we presented digital architecture, which significantly outperformed similar analogue systems and still exhibited competitive power efficiency. In this paper we extend this approach, introducing further improvements in processor functionality and performance, combined with more than 75% reduction in cell area and increase of fill-factor thanks to shifting the technology node from 0.35μm to 0.18μm. Current design (ASPA2, Figure 1) is based on the architecture used in our previous 19×22 ASPA1 chip [8].

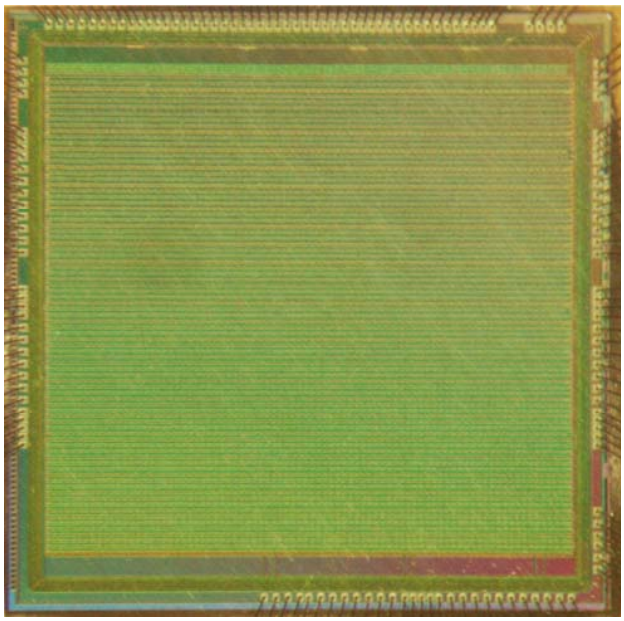

Figure 1: The ASPA2 chip microphotograph

Next section briefly describes the array architecture followed by the details of improved features of the processing cell. Then implementation details are described. The paper finishes with comparison results and conclusions.

II. Architecture

The architecture of the ASPA2 chip is presented in Figure 2. The pixel-parallel processing is facilitated by 80×80 array of locally interconnected processing elements (PE) mapped onto rectangular grid. Each PE combines a photo-sensor with one bit A/D converter and a simple digital microprocessor, which is comprised of eight 8-bit registers (two of which are shift-registers), eight 1-bit registers (which can be used as single 8-bit register), bit-serial ALU, flag register and communication unit responsible for multiplexing input data. There are two local data buses: Local Read Bus (LRB) – where the data is read into, and Local Write Bus (LWB) – where the data is read from. The array operates according to SIMD paradigm, while branching is provided by local flag indicator. Global operations are supported by a 'propagation

978-1-4244-5309-2/10/$26.00 ©2010 IEEE

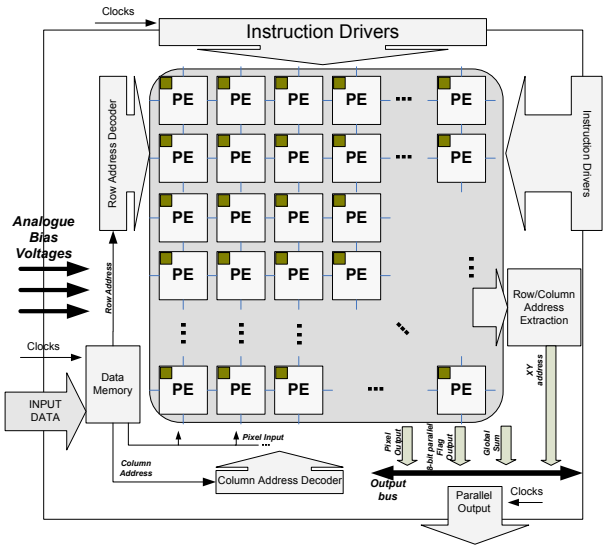

Figure 2: Block-diagram of the ASPA2

chain' [9] and executed asynchronously. The topology of the propagation network is controlled by local data in every PE.

All PE's execute instructions issued by the central controller (currently residing off chip) and distributed via periphery drivers on the chip. The data from pixels can be accessed in several ways. In addition to individual random pixel addressing, there is a possibility to address several pixels simultaneously. Furthermore, flexible read-out [8] facilitates parallel binary data output from 8 PE's at the same time as well as global summation and logic OR operation.

III. New Features

The processing capabilities of the overall vision chip are primarily defined by the functionality of the PE. Therefore the design of the PE cell is guided by the target applications. In our pursuit to realise a general-purpose device capable of efficient execution of a broad spectrum of low- and medium-level image processing applications, the PE contains a variety of arithmetic, logic and global processing units. Although the previous chip demonstrated high performance compared to other digital and analogue pixel-parallel architectures, we have identified a number of architectural shortcomings. In the presented work we attempted to improve certain operations, thus increasing the overall performance in most critical and frequently encountered tasks.

The majority of object reconstruction and tracking algorithms require a relatively large number of binary image representations and logic operations on those images. Hence every PE is required to support fast access to binary pixel descriptors as well as Boolean and conditional operations on this data. Therefore, in addition to eight dynamic (due to area optimization) 8-bit registers, intended for greyscale operations, every PE in ASPA2 incorporates eight individually addressable binary memory cells (Figure 3). Additionally, this memory can be accessed as a single 8-bit register (register *H*), expanding the local pixel memory to 72 bits. The schematic diagram of the binary memory is presented in Figure 4. The value of this register can either be transferred to the LRB as an 8-bit greyscale value, or passed as a single binary value to the flag register. The binary output is the result of a logic OR operation on bits, selected in the instruction word (*H flag*).

Additionally, each cell contains an 8-bit parallel inverter, linking LWB and LRB (see Figure 3). Because it is also bit-controlled, it enables inversion of only specified bits. Together with the OR operation in register *H*, it completes the Boolean basis so that any logic operation can be realised. Additional instant logic operations (XOR, NXOR) are executed by the ALU.

The output of register *H* serves as an input to the flag register. Such configuration enables fast (2 and 4 clock cycles) conditional operations based on the result of any logic function, e.g.:

if (A[i])
if (!B[j])
if (C[j] ***and*** D[k])
if (C[j] ***or*** D[k])
if (E[j] ***xor*** H[k])

Additional improvement has been made in terms of arithmetic operations. In previous implementation, the bit-serial ALU has been designed to realise only subtraction, so that addition can only be implemented by negating one term and then using it as the subtrahend. Such configuration led to inefficient use of additional memory, more complicated programming model and efficiency reduction. ASPA2 has a reconfigurable ALU, capable of executing both subtraction and addition. This has been achieved at the cost of only two

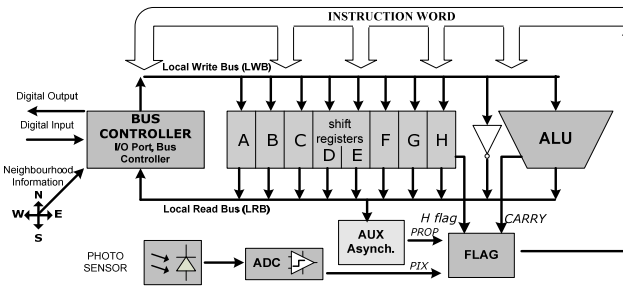

Figure 3: ASPA2 Processing Element

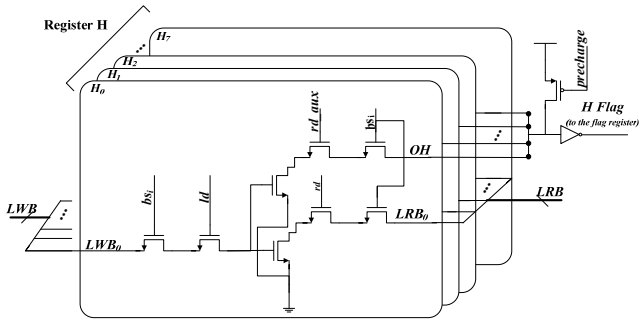

Figure 4: Dynamic register H with individually accessible bits.

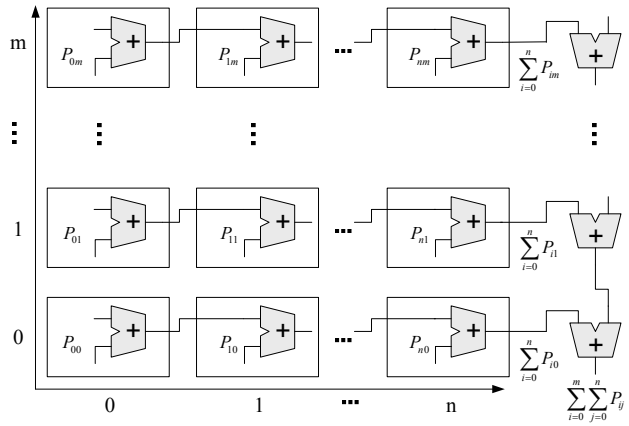

Figure 5: Chaining PE's ALUs for asynchronous global summation

additional transistors. More importantly, ALUs in neighbouring cells can now be chained together and configured as a single ripple-carry adder (Figure 5). This enables asynchronous row-wise global summation. The result of the global summation is stored in specially dedicated register at the periphery of the chip, whereas row-wise sub-sums are stored in the PEs in the most right column. Global addition is extensively used to gather image statistics, e.g. pixel count, histograms, etc.

IV. Implementation

One of the most critical aspects of the vision chip design is the size of the processing cell. In order to achieve high resolution at reasonable cost (physical chip size) the size of the cell has to be comparable with the pixel pitch of a regular CMOS image sensor, i.e. a few microns. Of course, size constraints have to be realistically weighted against required functionality.

ASPA2 have been implemented in a 0.18 μm 6-metal CMOS technology. The cell measures 51×54 μm^2 providing 363.1 cells/mm^2. The photo-sensor occupies 15×15 μm^2, hence the fill factor is 12.3%. For comparison, ASPA1 chip [8] had 100×117 μm^2 cell with fill factor ~3%. Thus the PE size reduction is 76%. It should be pointed out that such size shrink was achieved together with significant extensions of the PE's functional capabilities. Being comprised of 588 transistors it is 27% more "complex" than ASPA1 realisation. Significant contribution to the size optimisation has been achieved by the availability of additional routing layers, as anticipated in [8]. Only a small portion of the silicon space (less than 1%) is now occupied solely by routing, therefore further increase in the number of routing layers shouldn't result in significant size reduction. The microphotograph of the chip and layout comparison between ASPA1 and ASPA2 can be found in Figure 1 and Figure 6 respectively.

The current design also highlights technical challenges when designing large scale fine-grain SIMD processor arrays in a monolithic device, in particular signal integrity on instruction distribution wires. Being run through the entire array, these wires have significant load and coupling capacitance, resulting in unmatched skew and slew rate in different PEs. More importantly, accumulation of unmatched coupling capacitances (due both to the actual cell design and process variability) on different instruction lines results in a situation when these signals have different slack, which can cause incorrect operation and increased power consumption. This problem worsens with the increase of the operating frequency. A number of malfunctioning operations, caused by this problem, have been detected on ASPA2 and solved by introducing intermediate 'delay' instructions, thus slightly compromising performance in a few operations. This problem has been addressed in the next chip (ASPA2b, in fabrication) by more accurate matching of the periphery drivers.

Scaling the process technology also leads to reduction of the operating voltage. With V_{DD}=1.8V supply, operating at 75 MHz the 80×80 ASPA2 chip consumes 41.4 mW (worst case) thus resulting in 6.4 μW/cell.

V. Performance

The chip has been tested at 75 MHz. The general performance numbers are not much different from the ASPA1 version as the data flow remains very similar. At this frequency ASPA2 chip provides 240 GOPS (binary), 15.3 GOPS (8-bit greyscale) and 1.9 GOPS (8-bit greyscale products and quotients). Comprehensive post-layout simulations indicate that the chip is capable of running at 300MHz, so that performance results and power consumption are expected to quadruple. The debug system, which would support such operation, is under development and updated experimental results can be expected. The overall performance is significantly extended by the availability of new global operations, in particular global summation. Because every bit of the sum is calculated asynchronously across the array, the timing of this operation is almost unaffected by the operational frequency and mostly depends on the array size and operating voltage. Calculating single bit requires four microinstructions, three of which are synchronous and the forth one is asynchronous, propagating the sum along the rows and then column. Therefore the total time for bit calculations is equivalent to 3 clock cycles and signal propagation delay along the longest data path. For 80×80 ASPA2 chip at 1.8V the full 8-bit global sum calculation (worst case) needs 1.22 μs at 75MHz. This speed is sufficient to perform 819 global summations at 1000 frames per second, which is enough for the majority of vision applications. In addition to global sum, the array facilitates

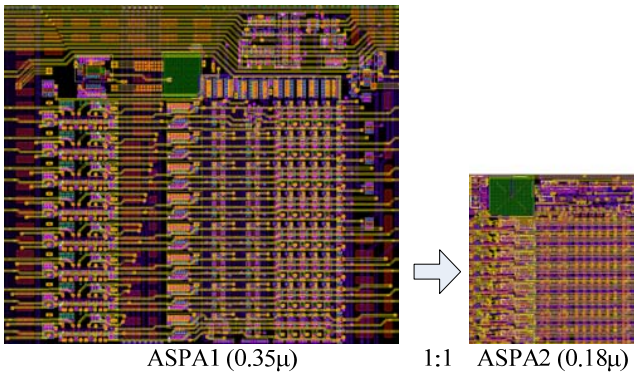

Figure 6: Scaling of the processing cell size.

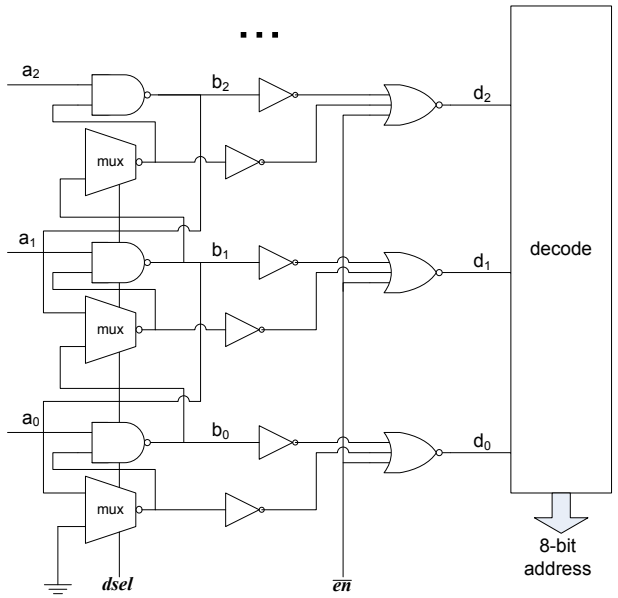

Figure 7: Asynchronous address extraction

asynchronous binary wave propagations (especially effective in morphological operations and segmentation). With less then 0.45 ns propagation delay per pixel, the global propagation takes less than 80 ns.

The functionality of the chip has also been improved by extending periphery operations, especially by asynchronous address extraction. This operation facilitates extraction of top, bottom, left and right object coordinates in asynchronous fashion. This functionality is provided by implementing the circuit, presented in Figure 7, on the right and at the bottom of the array. The binary inputs $a_0, a_1, \ldots, a_n$ represent the results of row- and column-wise NOR operation on binary flag outputs from selected cells. The output vector $(d_0, d_1, \ldots, d_n)$ has only one bit set to logic '1', which identifies the most right or left (top/bottom) segmented pixel, depending on the control signal ***dsel***. This vector is then decoded into numerical coordinate equivalent. The full operation of extracting 4 border coordinates of a single object (provided it is already segmented from the image) takes 214 ns at 75 MHz. In case of multiple objects – the extraction is performed sequentially on object by object basis.

Each cell requires 62 control signals and 3 analogue voltages. Instruction words are distributed horizontally and vertically, and are provided by the periphery drivers. At 75MHz ASPA2 provides 18 MPixel/second output stream via 8-bit global data-bus (the peak output is 2800 fps for 80×80 pixels without on-chip processing). A flexible addressing scheme allows selecting 8 pixels at the same time. Thus, if a binary output is required, the overall throughput increases to 150 MPixels/second. Considering the performance figures and output bandwidth, it should be noted that the integration time of the photo-sensor becomes the main limiting factor for high speed operations.

Since absolute performance figures per cell improved only slightly, compared to the previous version of the chip, the performance comparison to other chips (e.g. [2-4]) provided in [8] remains valid. However, as predicted, significant improvements have been achieved in terms of power and performance efficiency, areas which traditionally were dominated by analogue designs. Providing 373 GOPS/W and 871 MOPS/mm^2 ASPA2 compares favourably with analogue counterparts.

VI. Conclusions

The design of the 80×80 ASPA2 chip has been presented. The chip was fabricated in a 0.18 µm CMOS process. In addition to raw operating performance of 15.3 GOPS for greyscale operations its applicability is significantly increased by highly efficient global asynchronous operations.

ASPA2 chip represents an evolutionary work, based on our previous research. The main goal of this paper was to demonstrate how digital vision chip architecture can benefit from technology scaling. Simulations indicate that the design will also work at 90 nm node as well. Of course, deeper submicron technologies (45nm and smaller) will require additional solutions to counter high leakage currents in dynamic registers and stronger crosstalks. However, small retention time can be addressed, for example, by manual refreshing, which may incur relatively small performance penalty compared to the achieved benefits.

VII. Acknowledgement

This work has been supported by the EPSRC under grant no EP/D029759/1.

VIII. References

1. Moini, A., *Vision Chips or Seeing Silicon*. 1999: Kluwer Academic Publishers
2. Dudek, P., Hicks, P.J., *A general-purpose processor-per-pixel analogue SIMD vision chip.* IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 2005. **52**(1): p. 13-20.
3. Komuro, T., Kagami, S., and Ishikawa, M., *A dynamically reconfigurable SIMD processor for a vision chip.* IEEE Journal of Solid-State Circuits, 2004. **39**(1): p. 265-268.
4. Linan, G.C., et al., *A 1000 FPS at 128 x128 vision processor with 8-bit digitized I/O.* IEEE Journal of Solid-State Circuits, 2004. **39**(7): p. 1044-1055.
5. Dudek, P., *An asynchronous cellular logic network for trigger-wave image processing on fine-grain massively parallel arrays.* IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 2006. **53**(5): p. 354-358.
6. Galilee, B., et al., *Parallel asynchronous watershed algorithm-architecture.* IEEE Transactions on Parallel and Distributed Systems, 2007. **18**(1): p. 44-56.
7. Gies, V., Bernard, T.M., and Merigot, A. *Convergent micro-pipelines: a versatile operator for mixed asynchronous-synchronous computations*. in *IEEE International Symposium on Circuits and Systems (ISCAS)*. 2005. NJ, USA:p. 5242-5245.
8. Lopich, A. and Dudek, P. *ASPA: Focal plane digital processor array with asynchronous processing capabilities*. in *IEEE International Symposium on Circuits and Systems*. 2008. Seattle, USA:p. 1592-1595.
9. Lopich, A. and Dudek, P. *Implementation of an asynchronous cellular logic network as a co-processor for a general-purpose massively parallel array*. in *European Conference on Circuit Theory and Design (ECCTD '07)*. 2008. Seville, Spain:p. 84-87.