ASPA: Focal Plane Digital Processor Array with Asynchronous Processing Capabilities

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 implementation and experimental results for a digital vision chip that operates in mixed asynchronous/synchronous mode. Mixed configuration benefits from full programmability (discrete-time mode) and high operational performance in global image processing operations (continuous-time mode) thus extending the application field of smart sensors from low- to medium-level processing. A 19×22 proof-of-concept chip was fabricated and tested. At peak operational frequency (150MHz) each cell provides 9.6 MOPS thus achieving area utilization 820.8 MOPS/mm^2 and power efficiency 29 GOPS/W.**

I. Introduction

The idea of smart vision sensors or so-called 'vision chips', where sensing and processing is performed on a single silicon die, has been studied for more than a decade [1]. The main field of application for such devices has been low-level image processing, due to peculiar high degree of data parallelism and computational locality. Exploiting pixel-level parallelism and the fact that most of image preprocessing tasks is based on identical operations performed on every pixel, efficient architecture for a vision chip can be implemented as a fine-grain SIMD processor-per-pixel array. Various architectures have been previously described in the literature [2-7]. The majority of presented chips operate in SIMD mode, so that every cell executes the same instruction issued by a host controller in a synchronous manner. Yet, in image preprocessing there are a number of operations that are based on an iterative process, which results in a global data-flow across the pixel network. In such operations the result in each cell, despite processing only local data, implicitly depends on the entire array of pixels. Despite potential feasibility, it is apparent that synchronous SIMD architecture is not the most optimal tool for this kind of operations. Our attention was, therefore, drawn to a challenge of how to enable efficient execution of the global operations on a simple and compact parallel architecture. The successful solution would help to exploit the fine-grain parallelism more fully and significantly simplify the development of medium-level processing algorithms, thus extending application of smart sensors from preprocessing to more complex image analysis.

Another important issue associated with the design of smart vision sensors is the choice of the most appropriate (efficient) signal processing paradigm. Despite the higher power and area efficiency of analogue vision chips compared to their digital counterparts, the digital processing paradigm has more potential for scaling due to the continuous development of fabrication technologies. Robustness and noise free operation make digital circuits more favourable for building processor-per-pixel arrays. Presented design can outperform similar analogue systems and provide competitive power efficiency. This fact provides the motivation for further development of digital vision chips and suggests that such an approach will dominate in the design of massively parallel processor-per-pixel arrays.

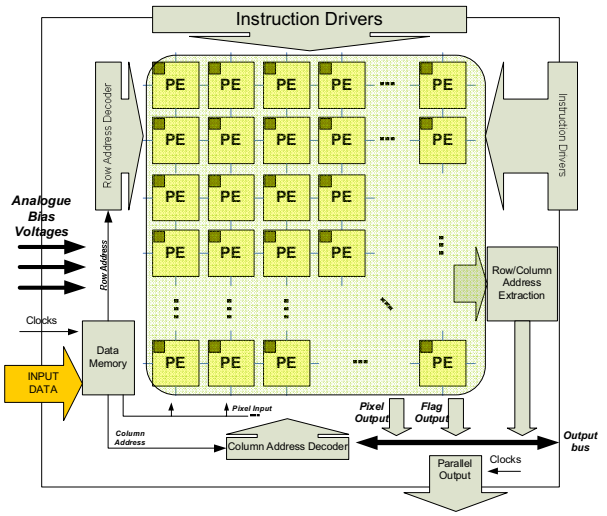

Figure 1: Architecture of the ASPA vision chip

Earlier we have proposed a novel architecture for the asynchronous/synchronous processor array (ASPA) [7]. In this paper we present the details of its hardware implementation, experimental results and performance evaluation. In the next section brief description of the architecture is provided, followed by details of hardware implementation. Then, performance characteristics of the chip are described and comparison with other image processing architectures is provided. The paper finishes with the design summary and conclusions.

II. Architecture

The description of the ASPA architecture can be found in [7] and here only brief details are provided. The ASPA chip can be represented as an array of locally interconnected processing cells placed on a 4-connected rectangular grid (Fig. 1). Processing cells execute the same instructions issued

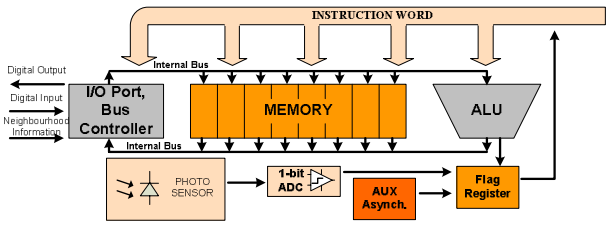

Figure 2: Block-diagram of the processing cell

by a central controller, however some degree of local autonomy is provided by a flag register.

Every processing element (PE) is connected to its four local neighbours, thus supporting local data transfers. The cell itself is a simple digital microprocessor, which consists of a bit-serial ALU, eight 8-bit general purpose registers (64 bits of memory per cell), photo-sensor, simple A/D converter, flag register and communication unit, responsible for multiplexing input data (Fig. 2). The cell operates as a data-path and all register-transfer operations are controlled by an instruction word.

In addition to aforementioned units, each cell contains a simple functional block called 'propagation chain', which facilitates global asynchronous wave-propagations across the entire pixel array. The topology of the propagation network is controlled by local data in every pixel. Produced binary trigger-waves are used to perform global operations such as object reconstruction and hole filling in a fast and power-efficient way [8].

In addition to conventional data processing the presented design allows chaining PEs into a single network, thus facilitating asynchronous distant data transfers, global feature extraction and data-driven pixel address extraction [9].

The major benefit of this approach is that global data-flow is handled in an asynchronous manner, thus pushing power consumption and performance characteristics to their absolute optimum.

III. Implementation

The first version of the ASPA chip has been implemented and tested. The chip microphotograph is shown in Fig. 3. The design is fabricated in a 0.35 µm CMOS technology; a 19×22 array occupies 9 mm^2. The PE circuitry consists of 460 transistors (most are of minimum size), 94% of them are n-type transistors (the design is based on dynamic logic). A simple photo-sensor circuit was incorporated within each pixel with sensor occupying 11.75×15.7 μm^2. The area of each PE is 100×117μm^2, thus achieving 85.5 cells/mm^2. Hence, a 128×128 array could be fabricated on a 200 mm^2 chip (including peripheral circuitry). The overall performance is estimated according to the executed operation (binary, greyscale, unsigned multiplication and division) and operation mode (synchronous, asynchronous). The chip has been tested with a 37.5 MHz clock (the maximum frequency of the available pattern generator). However comprehensive post-layout simulation results confirm the correct operation at frequencies up to 150 MHz. At 37.5 MHz operational clock every PE in the ASPA provides 37.5 MOPS (binary), 2.4 MOPS (greyscale) and 0.3 MOPS (for unsigned products and quotients) in a synchronous mode. The achieved area usage parameter is 205.2 MOPS/mm^2 (for greyscale operations). At 150 MHz clock these figures increase in four times. Interpolated performance characteristics for a 128×128 array are as follows: 0.61 TOPS for binary operations, 39.3 GOPS for greyscale operations and 4.9 GOPS for unsigned products. Operating at 150 MHz clock the chip consumes ~350 µW/cell. Operating at 37.5 MHz clock the 19×22 array consumed only 26.35 mW (V_{DD}=2.5V), which makes it 63.04 µW/cell.

Furthermore, the overall performance can be significantly boosted by the gain from asynchronous operations. With 0.76 ns propagation delay and the maximum propagation length 128 pixels, the equivalent performance (corresponding to a synchronous SIMD 128×128 array) during asynchronous processing is 168 TOPS. Of course, this estimation does not reflect the effective performance, however, the fact that the global operation takes only a few clock cycles to be accomplished reduces the processing time significantly. In case of trigger-wave propagations on the ASPA, the worst case energy consumption (i.e. all the PEs lay within the propagation space and are triggered) is 0.4 pJ per PE.

IV. Image Processing

As the ASPA chip is a fully programmable general-purpose processor, a wide range of image processing algorithms can be implemented. In order to verify the processing, images were loaded into the chip through an input interface. Subsequently the processing output results were compared with numeric simulations. The comparison indicated that the chip operates exactly according to specifications. Fig. 4 and Fig. 5 demonstrate examples of different image processing tasks executed on ASPA. Sobel edge detection, smoothing and median filtering are performed on grey-scale images in 3×3 neighbourhood. Significant number of useful morphological image processing algorithms deals with images represented in binary format (binary

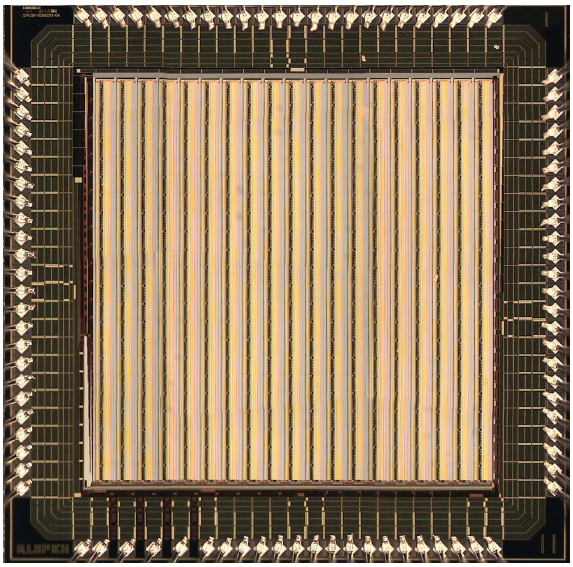

Figure 3: ASPA Microphotograph

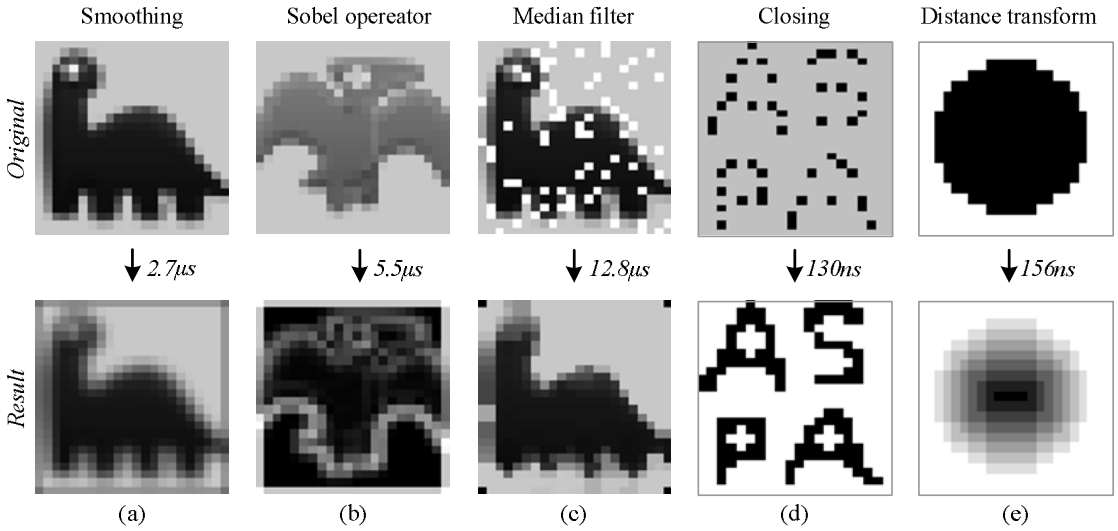

Figure 4: Results of image processing on the ASPA vision chip with operation times at 37.5 MHz clock. When operated at 150 MHz, times for figures (a)-(d) are reduced in four times and the distance transform takes 42 ns.

dilation, erosion, skeletonization). As can be noticed, the ASPA can execute many of these tasks in an efficient way, as some of the logic operations are performed on data buses during data transfers, i.e. at no additional cost. The implementation example of *closing* operation, where sequential dilation and erosion helps to restore object's connectivity, is presented in Fig. 4(d). By utilising the features of the bus controller and shift registers it is possible to perform simple processing while transferring data between cells. In particular, it is possible to perform a simplified version of the minimum and increment operation [9]. Thus, a distance map can be computed in an asynchronous manner. An example of using binary trigger-waves for geodesic reconstruction is presented in Fig. 5. With the propagation delay of 0.76 ns per pixel the equivalent synchronous implementation on ASPA in iterative manner would require operation at 2.7 GHz (two operations per cell), providing over 44.2×10^{12} operations per second on a 128×128 array. Such performance corresponds to a maximum throughput of 10.3×10^{6} fps. In addition to aforementioned operations, there is a facility to perform data-driven address extraction. Locating a border (left or bottom) of the segmented object takes only 9 operations or 234 ns at 37.5 MHz clock.

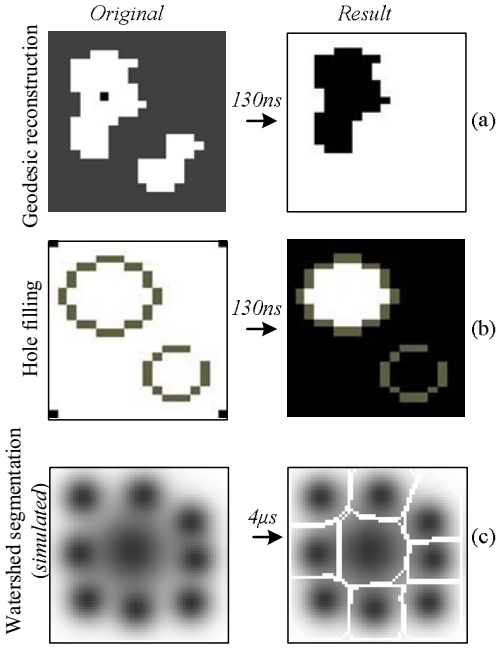

Figure 5: Global operations on the ASPA chip.

Achieved results indicate that certain global operations can be used as basic operations with small processing time to implement computationally expensive algorithms. In other words, complex algorithms can be decomposed into a set of global and local operations that are feasible on the relatively simple ASPA architecture. For example, the watershed transformation illustrated in Fig. 5(c) for a 64×64 array would be achieved within 4 μs at 150 MHz operational clock.

V. Comparison with other vision chips

Absolute figures of merit (such as "operations per second"), often produced in literature, can be misleading as they do not reflect the performance in 'real' applications. It is important to realize that different designs can exhibit different performance characteristics depending on the target application and algorithm implementation. Therefore, performance estimation techniques provide more sensible information when they are based on measuring the performance of the device on a fixed set of tasks. Available benchmarks for image analysis and filtering operations allowed us to compare the ASPA chip with top-of-the-line DSP TMS320C6455 [10]. The results are presented in Table 1. It can be noticed, that the ASPA vision chip significantly outperforms the DSP at low-level image preprocessing despite operating at much lower frequencies.

The result of comparison of the fabricated chip and the expected characteristics of the scaled 128×128 ASPA chip with its closest digital and analogue counterparts are summarized in Table 2. The overall performance is estimated

Table 1: Comparison of image preprocessing on TI DSP and ASPA scaled for 128×128 array

Benchmark	Total time, ns TMS320DM6455	ASPA total time @37.5MHz, ns	ASPA total time @ 150MHz, ns
Boundary Structural Operator	21008	416	48
3x3 Binary Dilation	2192	130	15
3x3 Binary Erosion	2192	130	15
Perimeter Structural Operator	2160	130	15
Sobel Edge Detection	17264	5590	645
Image Thresholding	2368	468	54
3x3 Median Filter	12298	12740	1470

for grey-scale operations, by which we imply arithmetic operations (excluding multiplication) of maximum precision for analogue designs and 8-bit for digital. As it can be seen from the table, according to reported data, ACE16k appears to outperform other designs in terms of power and area efficiency by a significant margin. However, it is important to understand that this chip is positioned as a 'CNN-inspired' SIMD array and the performance is estimated from the CNN computational model perspective. Indeed, despite providing 330 GOPS performance, simple convolution operations such as smoothing and Sobel edge detection are reported to take approximately 4μs (with required calibrations) [11]. At the same time, the ASPA chip with 2.4 MOPS/cell executes these operations in 2.7μs and 5.5μs correspondingly (at 37.5 MHz).

The results indicate that the computational power of the array is suitable for real-time applications and compares favourably with other digital and analogue vision chips. The achieved power efficiency is comparable with other digital vision chips but yields to analogue counterparts. However, this parameter is expected to improve when the design is fabricated in a scaled down technology.

The performance of the ASPA chip during asynchronous wave propagations can also be favourably compared with arrays that provide similar facilities [3, 12].

VI. Conclusions

The presented design attempts to combine synchronous and asynchronous approaches in order to extend the range of its applications from low- to medium-level image processing. The 22×19 ASPA chip, which serves as a proof-of-concept design, has been fabricated and tested. It demonstrates efficient execution of global operations, implemented in asynchronous manner. Designed in a 4-metal 0.35 μm CMOS technology each processing cell occupies 100×117 μm^2 area. Our recent designs indicate that utilization of two more routing layers (6 in total) reduces the cell area by 30%. Operating at 2.5V and 37.5 MHz the chip delivers 15.6 GOPS for binary and 1 GOPS for grey-scale operations. At 3.3V and 150 MHz these figures are quadrupled.

This chip compares favourably with other digital and analogue architectures that are based on a similar processor-per-pixel approach. Although some analogue vision chips are more power efficient, fully digital approach is suitable for the design of compact massively parallel architecture, and it promises relatively simple adaptation to finer technologies.

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] Eklund, J.-E., Svensson,C., Astrom, A., *VLSI Implementation of a Focal Plane Image Processor-A Realization of the Near-Sensor Image Processing Concept.* IEEE Tr. on VLSI Systems,1996.**4**(3):p.322-335.

[4] Komuro, T., Ishii, I., Ishikawa, M.,Yoshida, A., *A digital vision chip specialized for high-speed target tracking.* IEEE Transactions on Electron Devices, 2003. **50**(1): p. 191-199.

[5] Linan, G., Espejo, S., Dominguez-Castro, R., Rodriguez-Vazquez, A., *Architectural and Basic Circuit Considerations for a Flexible 128x128 Mixed-Signal SIMD Vision Chip.* Analog Integrated Circuits and Signal Processing, 2002. **33**: p. 179-190.

[6] Paillet, F., Mercier, D.,Bernard, T.M. *Second generation programmable artificial retina*. in *Twelfth Annual IEEE International ASIC/SOC Conference, 15-18 Sept. 1999*. USA: IEEE. p.304-309

[7] Lopich, A., Dudek, P. *Architecture of a VLSI cellular processor array for synchronous/asynchronous image processing*. in *IEEE ISCAS*. 2006. p.3618-3621

[8] Lopich, A., 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*. 2007. Seville, Spain: IEEE. p.84-87

[9] Lopich, A., Dudek P. *Global operations in SIMD cellular processor arrays employing functional asynchronism*. in Sixth Workshop on Computer Architecture for Machine Perception *(CAMPS)2006*. p.18-23

[10] *Texas Instruments DSP TMS320C6455-1000.* http://focus.ti.com/lit/ds/symlink/tms320c6455.pdf, May, 2007.

[11] Cembrano, G.L., Rodriguez-Vazquez, A., Galan, et.al. R., *A 1000 FPS at 128x128 vision processor with 8-bit digitized I/O.* IEEE Journal of Solid-State Circuits, 2004. **39**(7): p. 1044-1055.

[12] Flak,J.,Laiho,M.,Paasio,A.,Halonen, K., *Dense CMOS implementation of a binary-programmable cellular neural network.* Int. Journal of Circuit Theory and Applications, 2006. **34**(4): p. 429-443.

Table 2: Comparison of analogue and digital vision chips

Reference	**PVLSAR 2.2 [6]**	**NSIP [3]**	**SPE [4]**	**ACE16k [5]**	**SCAMP [2]**	**ASPA @37.5MHz**	**ASPA @150MHz**
Array Size	*128×128*	*32×32*	*64×64*	*128×128*	*128×128*	*19×22*	*128×128*
Cell Size, μm^2	*60×60*	*118x118*	*67.4×67.4*	*75.7×77.3*	*49.35×49.35*	*100×117*	*100×117*
Technology, μm	*0.8*	*0.8*	*0.35*	*0.35*	*0.35*	*0.35*	*0.35*
Processing type	*digital*	*digital*	*digital*	*analogue*	*analogue*	*digital*	*digital*
Memory per cell	*5 bit*	*8 bit*	*24 bit*	*8 an. reg,4 Bin*	*9 an. reg.*	*64 bit*	*64 bit*
Performance, GOPS	*49*	*1.9*	*6.4*	*330*	*20*	*1*	*157*
Die Size, mm^2	*76*	*25*	*29.2*	*145.2*	*50*	*9*	*213.5*
Power per chip	*1 W*	*0.1W*	*N/A*	*2.9W*	*240mW*	*26.4mW*	*5.4W*
P_A, MOPS/mm^2	*92.6*	*1.1*	*343*	*3800*	*512*	*205.2*	*820.8*
P_E, GOPS/W	*5.4*	*25.2*	*N/A*	*180*	*85.3*	*38*	*29*