A 39x48 GENERAL-PURPOSE FOCAL-PLANE PROCESSOR ARRAY INTEGRATED CIRCUIT

Piotr Dudek

Department of Electrical Engineering and Electronics
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester M60 1QD, United Kingdom
p.dudek@umist.ac.uk

ABSTRACT

This paper presents the implementation of a general-purpose programmable vision chip, with a 39×48 SIMD processor-per-pixel array, fabricated in a 0.35µm CMOS technology. The chip employs Analogue Processing Elements to achieve cell density of 410 cells/mm^2. The array operates at 1.25MHz with power consumption of 12µW/cell and executes low-level image-processing algorithms in real-time. Chip architecture, circuit and layout design issues are discussed. Experimental results are presented.

1. INTRODUCTION

In a retina-like smart sensor thousands of tiny processors have to be placed, together with photodetectors, on a single integrated circuit. The trade-offs between speed, accuracy, functionality, processor area and power consumption have to be very carefully resolved. The choices of a suitable architecture and efficient circuit techniques are of critical importance. Digital vision chips, based on bit-serial SIMD (Single Instruction Multiple Data) processor arrays have been reported in the literature [1-2]. Compact digital processing elements can be built, but due to limited local memory present in these devices they are suitable mostly for binary image processing. However, some gray-scale image processing is usually required and techniques based on analogue circuits seem to be more suitable for this purpose than digital solutions (at least in present-day CMOS technologies). The analogue approaches appear to offer better efficiency in terms of speed, area and power consumption [3-4].

Previously, we proposed an approach to the design of a general-purpose programmable vision chip, which is essentially a combination of a "conventional" SIMD architecture and an "unconventional" processing element circuitry [4]. A new silicon implementation of our vision chip concept (i.e. the SCAMP-2 chip) has been recently fabricated. In the new chip the design of the Analogue Processing Element (APE) has been further simplified and refined [5], while the Flexible Global Readout Architecture [6] has been added, to enhance chip's ability to perform data reduction at the sensor level. In this paper the chip design is overviewed and test results are presented.

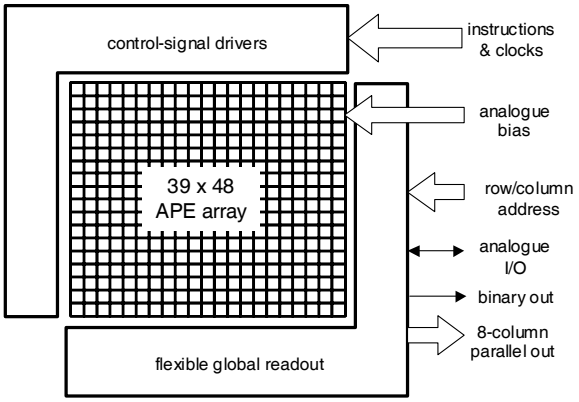

Fig.1. Architecture of the SCAMP-2 vision chip

2. CHIP ARCHITECTURE

The processing core of the chip is formed by a 2-D mesh-connected processor-per-pixel array, shown in Figure 1. The processing elements in the array (APEs) are implemented as "analogue microprocessors" (they process data according to a software program, but achieve this using analogue circuits, and store data in a form of analogue sampled signals). A single controller issues instructions, which are broadcast to each APE in the array, and the APEs execute instructions in parallel, according to the SIMD paradigm, each operating on their local data. The pixel-parallel arrangement is perfectly suited to execute early vision (low-level image processing) algorithms.

The design of the APE in the SCAMP-2 chip has been previously reported [5]. Its architecture is shown in Figure 2. Briefly, each APE contains a photodetector, nine "registers" (each capable of storing analogue data, e.g. a gray-scale pixel value or another scalar variable), a 4-neighbour local communication network, I/O circuits for external data transfers, and a local-autonomy activity-flag circuit.

The overall principle, when designing the APE, was that of a "minimum hardware" processor. It was motivated by the desire to reduce the size of the APE, so that large resolutions of pixel-per-processor arrays are feasible. At the same time, though, the accuracy of processing had to be maintained at an acceptable level. A further consideration was the reduction in power consumption.

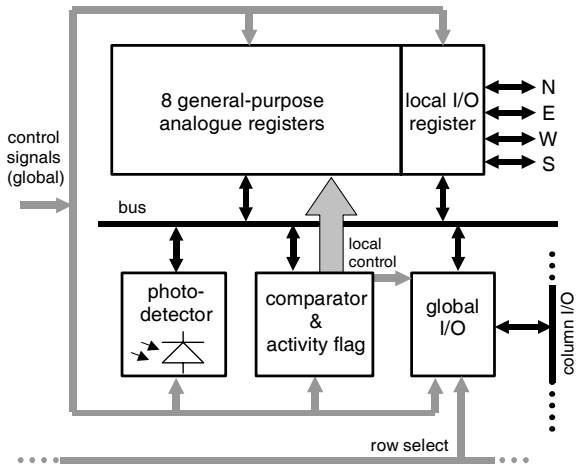

Fig.2. Architecture of the APE

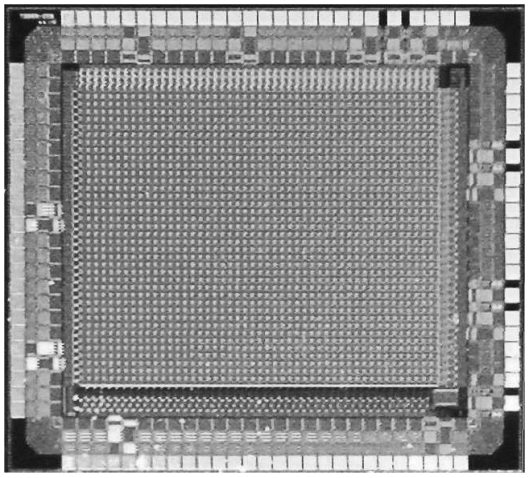

Fig.3. Chip microphotograph

2.1. ALU-free design

A primary function of any algorithmically-programmable processor (i.e. a Universal Turing Machine) is to perform a programmed sequence of arithmetic and logic operations on data stored in the memory. On each machine instruction cycle the arguments of an operation are taken from the memory (e.g. processor's registers), presented to inputs of an appropriately configured computational circuit (known to microprocessor designers as ALU - Arithmetic Logic Unit), and the result of the operation is stored back into the memory.

The design of the APE follows this principle. However, it is notable that no dedicated hardware exists to implement arithmetic operations. Instead, the basic operations of addition, inversion and division are executed directly in the registers/analogue bus system, by means of switched-current signal processing techniques [5]. This results in significant silicon area savings (no need for ALU hardware). Consequently, it can also improve the accuracy of processing - increasing the size of transistors in the registers (in the present implementation, the analogue registers occupy almost 70% of the APE area) reduces the errors associated with analogue storage, and at the same time improves the transistor matching required, for example, to achieve accurate multiplication coefficients for convolution kernels, etc.

It has to be noted, that in this "minimum hardware" APE, the silicon area and accuracy improvements are achieved at a cost of a simplified instruction set and therefore somewhat reduced processing speeds. However, since the typical low-level image processing algorithms are rather simple, this massively parallel system easily achieves performance adequate for real-time computer vision.

3. IMPLEMENTATION

The chip has been implemented in a 0.35µm single-poly 3-metal layer CMOS technology. The chip microphotograph is shown in Figure 3. The 39×48 processor array is surrounded by the control-signal distribution network and the read-out circuitry. The total chip area, including I/O pads, is equal to 10mm^2. (The chip is a scaled-down version of a planned 128×128 array). The APE cell size is equal to 49.35µm×49.35µm. The array operates at 1.25MHz clock, with 2.5V supply voltage and maximum power consumption of 12µW per cell.

3.1. Control signals

Control signals, which are used to configure APEs in the array, are derived from instruction-code-words that have to be provided to the chip on each clock cycle. This means, that the controller (sequencer), issuing instructions to the SIMD array, is located off-chip. This is a preferred solution for a proof-of-concept research design, as a suitable controller can be easily implemented on a programmable-logic device (or replaced by lab instruments as required). Similarly, all analogue bias voltages are provided externally.

Control signals are routed horizontally and vertically, over the processor array. A complete set of control-signal drivers exists for each row/column of the array. Global analogue biasing voltages (all of them driving high-impedance nodes) are routed along other control signals. Great care has been taken during the layout stage to ensure that the coupling between digital control signals and sensitive analogue circuit nodes is minimized where necessary. It has to be said, that using a 3-metal layer technology is somewhat restrictive, especially when it comes to providing good (low impedance) power supply rails, and while the results have been satisfactory for this implementation it is expected that the availability of another metal layer would further improve the noise performance of the circuits (especially if the array size is scaled up).

3.2. Read-out

The peripherial circuitry includes the Flexible Global Readout Architecture, described in more detail in [6]. It permits addressing individual APEs in the array (for random-access readout) as well as groups of the APEs, to facilitate certain global operations. It is expected that in typical high-frame-rate applications reading-out of full gray-scale images will not be necessary. Instead, some image "descriptors" or features will be obtained from the sensor/processor array system (e.g. coordinates of targets, counts of objects, etc.). Nevertheless, a mechanism for column-parallel read-out of binary images (i.e. processing results, such as an edge map, etc.) is also provided, via an 8-bit output port. An analogue read-out of gray-scale images (or data arrays corresponding to processing results) is also possible. The chip also accepts an analogue input data. There are no on-chip A/D or D/A converters in the present implementation. Instead, off-the-shelf components are used in the test system. Again, this option has been chosen to simplify the prototype and allowed to concentrate the chip design effort on the more innovative parts of the system.

3.3. Image Sensor

An image sensor is embedded within the processor array on a pixel-per-processor basis. An n-diffusion diode is used as a photodetector. The pixel fill-factor is approximately 5.6%. The design of the image sensor is fairly simple and has not been optimized for optical performance. However, care has been taken to minimize the mismatch by appropriately scaling the transistors. The measured fixed-pattern noise is equal to 0.4% (rms). The photodetector can operate in the integration mode – the integration capacitor is formed by the inherent capacitance of the photodiode and the gate capacitance of an n-MOS transistor, which is used (in a cascode configuration) to provide voltage-to-current conversion. The output current values are matched to the analogue register range (0 to 1.7μA). The sensitivity of the photodetector can be regulated, by changing the reset voltage and the transconductance of the transistor stage. The light-to-current characteristic can be close-to-linear for dark pixels and compressed (quadratically) for bright pixels. Additionally, the photodetector can operate in a logarithmic-compression, continuous mode.

The metal layers used for routing of control-signals and power supply voltages are also used as an optical shield, to reduce the incident light in the processing part of the APE, while providing an opening over the photodiode area to maximise the light sensitivity of the photosensor. As a result, the measured light-induced leakage in the analogue registers is 50 times below the sensitivity of the photodetector (i.e. if full-scale pixel brightness value is obtained after 10 ms integration time, for instance, then during this time the analogue value held in a register will change by 2% of the maximum value). This is an acceptable value, particularly since usually the integration time is much longer than the required retention time of an analogue value. However, the leakage error has to be taken into account if data is to be retained for longer periods (e.g. a few video frames). Furthermore, it has to be noted that just like any other focal-plane processor array, the chip could be "blinded" by a strong light source, leading to unpredictable processing results. This fact should be detected by the vision chip system, if it is required to work in an uncontrolled environment.

4. ACCURACY

The registers in the APE are implemented as S^2I memory cells (with additional features allowing to conditionally disable current-storage operation as well as switch-off biasing currents to save power when the registers are not being accessed). The storage of data in analogue memories, however, is not error-free. Apart from the fact that leakage currents lead to a change in the value of the stored sample, as described above, there are also errors associated with the sampling process (clock feedthrough and charge injection from MOS switches, output conductance effects) as well as noise. The arithmetic operations performed using analogue circuits are not perfectly accurate either. Furthermore, these errors are accumulated as the data is repeatedly sampled and processed throughout the execution of an algorithm. It is therefore of paramount importance, that the analogue errors are kept at an acceptable level.

The measured error of the single transfer operation (i.e. loading one register with the analogue value stored in the other register) reveals a signal-independent error of 56nA, which amounts to 3.3% of the maximum signal level of 1.7μA. However, it can be shown [4] that this error can be easily compensated algorithmically (each storage performs signal inversion, thus after two operations the offset error cancels out). More important is the signal-dependent error of the transfer operation. Its magnitude is equal to 6nA (i.e. 0.35%). Register-to-registers variations of the error, which are the most important as they will determine the fixed-pattern noise in the processing result, are equal to 0.042% (rms). Finally, there is a random noise associated with each transfer operation, and this was measured to be equal to 0.092% (rms).

It is not sufficient, though, to consider the accuracy of analogue memories only. The accuracy of arithmetic operations is also very important. An addition operation relies on current summation and is accurate, apart from a small error (similar to the transfer operation error), which can be attributed to the limited accuracy of the analogue memory cells. The division operation, on the

other hand, is inaccurate, as it relies on matching between the transistors in two or more registers. In the APE, the fixed-pattern noise of a straightforward division-by-two instruction (performed by dividing a current between two memory cells) has been measured to be equal to 2.32%. However, this can be significantly reduced, to below 0.2%, with a five-step algorithm based on an error compensation technique described in [7]. Overall, it is expected that the achieved levels of accuracy should be sufficient for robust implementation of a majority of low-level image processing algorithms. It also should be noted, that due to the "dividing in registers" concept, the mismatch problems associated with our approach are inherently smaller than ones present in CNN-based processors [3] (which require 9 multipliers, in addition to memory cells, to be fitted into the limited processor area).

5. IMAGE PROCESSING EXAMPLES

The SCAMP-2 architecture is that of a software-programmable general-purpose processor array for which a variety of image processing algorithms can be developed. Figure 4 includes examples of two simple algorithms, which have been implemented and executed on the SCAMP chip. The edge detection result is the sum of the absolute values of two images, each obtained by convolving the input image with a 3×3 Sobel convolution kernel (horizontal and vertical). The median filtering is performed in a 3×3 neighbourhood. For comparison, results of the same algorithms in an ideal case (calculated from the input images using numerical computation) are shown in Figure 4. The overall error is due to the accumulation of analogue processing errors and noise. The total rms difference between experimental and ideal results (excluding borders) is 2.1% for edge detection and 1.2% for median filtering. The images were obtained at 25 frames per second (fps) with processor array operating with a 1.25MHz clock. The processing times for the two algorithms are as follows: 30.3μs for edge detection and 157μs for median filtering. This, potentially, enables operation at several thousands fps. It is also worth noting, that the total power consumption in the processor array operating at 25 fps is only 73nW per pixel for the edge detection algorithm and 380nW per pixel for the median filter algorithm. To the best of our knowledge these power consumption figures are lower for our chip than can be achieved by any other processor.

6. CONCLUSIONS

A general-purpose programmable vision chip with a 39×48 SIMD processor-per-pixel array has been designed, fabricated, and tested. The concept of a "minimum hardware analogue microprocessor" allowed the size of the APE to be reduced so that cell density of 410 cells/mm^2 was achieved. At the same time, good accuracy of processing (particularly in terms of fixed-pattern-noise) has been achieved. The speed of the processing is appropriate to execute early vision algorithms in high frame-rate applications, while the extremely low power consumption makes the chip suitable for low power applications.

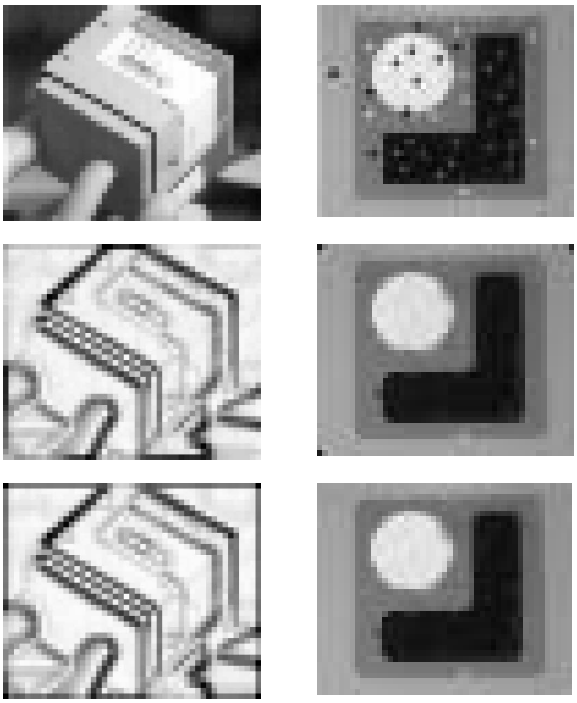

Figure 4. Results of image processing on SCAMP-2: left: edge detection, right: median filter. top: original images; middle: results obtained on chip, bottom: comparison with ideal (numerical) processing results.

ACKNOWLEDGEMENT

This work has been supported by the EPSRC under grant no. GR/R52688/01

REFERENCES

[1] M.Ishikawa et. al. "A CMOS Vision Chip with SIMD Processing Element Array for 1ms Image Processing", Proc. ISSCC'99, TP 12.2, 1999.

[2] F. Paillet, et. al., "Making the most of 15kλ^2 silicon area for a digital retina", Proc. SPIE, vol. 3410, AFPAEC'98. 1998.

[3] G. Liñán-Cembrano et al. "A Processing Element Architecture for High-Density Focal-Plane Analog Programmable Array Processor Arrays", Proc. ISCAS 2002, vol.III, pp.341-344, May 2002.

[4] P.Dudek and P.J.Hicks, "An Analogue SIMD Focal Plane Processor Array", IEEE International Symposium on Circuits and Systems, ISCAS 2001, vol.IV, pp.490-493, May 2001.

[5] P.Dudek, "A Processing Element for an Analogue SIMD Vision Chip", Proc. European Conference on Circuit Theory and Design, ECCTD'03, vol.III, pp.221-224, September 2003.

[6] P.Dudek, "A Flexible Global Readout Architecture for an Analogue SIMD Vision Chip", Proc. ISCAS 2003, May 2003.

[7] J.-S. Wang and C.-L. Wey, "Accurate CMOS Switched-Current Divider Circuits", Proc. ISCAS'98, vol I, pp.53-56, May 1998.