A pixel-parallel cellular processor array in a stacked three-layer 3D silicon-on-insulator technology

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

Viktor Gruev
Department of Computer Science and Engineering
Washington University in St.Louis
Missouri, USA
vgruev@cse.wustl.edu

***Abstract*—This paper presents the design of a vertically-integrated image sensor/processor device, implemented in a fully stacked 3-layer three-dimensional (3D) silicon on insulator (SOI) 150nm CMOS technology. This prototype 'vision chip' contains a 32×32 pixel-parallel processor array. Three silicon layers contain current-mode image sensors, current-mode analogue circuits and digital logic circuits, respectively. The two bottom layers form a mixed-mode cellular processor array, which operates in SIMD mode, and processes the image data acquired by the top-layer backside illuminated photosensor circuit. The intra-processor inter-layer communication is achieved by means of through-silicon vias, and the system is partitioned to minimise the area overhead associated with this communication. The processor comprises 4 analogue and 12 binary registers, and supports arithmetic and logic operations. Various sensor structures have been implemented to evaluate the efficiency of photo-sensing in SOI technology. The prototype circuit measures 2mm×2mm, with 30µm×30µm pixel pitch. The architecture and circuit design issues are presented in the paper.**

I. INTRODUCTION

A system that integrates image sensing and image processing on a single microelectronic circuit is known as a 'vision chip' [1]. Such a device works as an image sensor, acquiring images projected onto its surface through a lens, using photodetector circuitry. However, unlike a conventional image sensor which simply outputs the images to external devices for further processing, a vision chip contains processing circuits right next to the sensors. In the most advanced forms, these processing circuits are fully functional, albeit simple, "microprocessors", providing programmable image processing system. In our earlier work [2-5] we have developed a number of devices based on such pixel-per-processor architecture, using various approaches to the processing element design, including analogue, and digital synchronous/asynchronous processing cells. Several other researchers have proposed pixel-per-processor systems, integrating sensors, analogue and digital processing circuits as well, including pioneering work by Bernard et al. [6] and silicon implementations [7, 8] inspired by the CNN Universal Machine concept [9]. There are also numerous implementation of sensor/processor integrated systems, that offer a somewhat lesser degree of programmability, see [1] for a review of early developments.

The integration of processing right next to the sensors offers several distinct benefits, improving sensor-processor bandwidth (enabling very high sensing rate applications), maximizing performance through fully exploiting massive parallelism of low-level image processing tasks and integrating thousands of processors in a single chip, reducing power consumption, as most data is processed and transmitted only locally. There are, however, some shortcomings of this technique. As processors are placed next to the sensors the fill-factor (ratio of photosensitive area to the total pixel area) is very significantly reduced, leading to lower sensitivity of the device. Furthermore, it is not optimal to use the same fabrication technology for CMOS image sensing and processing circuitry. The achievable resolutions are also severely constrained, as pixel size must accommodate a reasonably complex processing circuitry, leading to pixel area of hundreds of μm^2 in today's technologies (as compared with only a few μm^2 needed for a pixel in a plain image sensor). These limitations have led some researchers to seek alternatives to processor-per-pixel integration in order to design practical image pre-processing devices, for example using linear [10] or coarse-grained [11] processor array architectures, lowering degree of parallelism, and trading-off potential performance & power benefits of fully pixel-parallel solutions.

However, new silicon integration technologies, based on 3D wafer stacking using through silicon vias (TSV) [12, 13], are promising to solve most of the problems associated with the planar 2D pixel-processor approach, enabling an elegant solution to sensor/processor integration. A sensor array, with optimized noise and sensitivity, can be placed on top of CMOS processing circuitry. Such devices have already been demonstrated [14] and it is expected that TSV technology will rapidly mature yielding commercial applications in the area of CMOS image sensors, followed by more general applications. There have been also a number of pioneering designs that integrate in a 3D structure image sensors and more elaborate processing circuitry [15, 16].

In this paper we describe a vision chip with a general-purpose software programmable pixel-parallel processor array. We present the design of a prototype 3D integrated circuit, which integrates sensing, analogue, and digital processing, implemented in a 3-layer TSV 150 nm SOI CMOS technology

978-1-4244-3896-9/09/$25.00 ©2009 IEEE

from MIT LL. Our earlier work [2-5] has demonstrated the basic components of the presented approach. Here, we show how the architecture and circuitry has been adapted to the 3D SOI technology.

II. System Architecture

The developed device follows the basic outline of the system shown in Figure 1. Three silicon tiers have been organised as sensing, analogue and digital processing layers. These are functionally integrated into a single sensor/processor array, so that it is convenient to think of a single 3D processing element (PE) 'cell', spanning across all three layers. These PE cells are then arranged on a 2D grid, with nearest-neighbour connections.

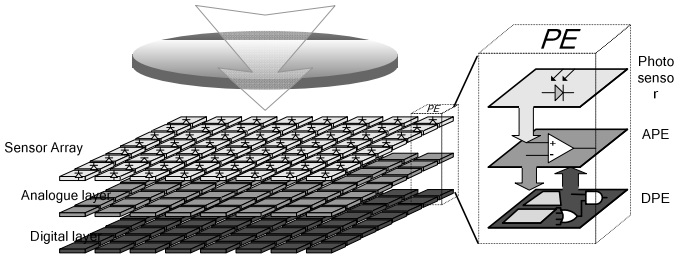

Figure 1. Overview of the proposed vision chip. Individual sensing/ processing elements (PEs) are physically distributed across 3 silicon tiers.

Overall, the system follows the SIMD (Single Instruction stream Multiple Data streams) pixel-parallel approach to the vision chip design, that we used in our earlier devices [3,4]. Acquired pixel data is processed according to instruction-level program, which is broadcast to all PEs in the array from a single controller. The controller has not been integrated onto this chip, but will be implemented externally (this is, of course, only a pragmatic approach for the research prototype, and it can be easily seen that a full controller, memory, interfaces, etc. could be integrated in a complete system-on-a-chip solution). The PEs all perform the same instruction on their respective data, but some degree of autonomy is achieved through the use of activity flags, which selectively enable/disable individual PEs in the array, based on the result of previous operations on local data. We have implemented both analogue and digital processing circuits, each providing optimised performance for specific tasks. The analogue processors have superior efficiency when performing gray-scale operations with moderate accuracy, while digital processors offer increased robustness to noise, higher speed, and more compact design for binary operations, long-term storage, or high-precision computations. The processor is distributed across two layers, with middle layer containing analogue processors and bottom layer containing digital processing circuits. The top layer is responsible for image acquisition, and consists of current-mode photodetector circuits.

Functionally, the system has the following capabilities:

- Sensor layer: photodetectors work in integration mode. They can be reset (globally) and the value of each sensor can be accessed by the analogue processor.
- Analogue layer: provides gray-level (continuous value) short-term storage and arithmetic operations of addition, subtraction, division by two, and comparison. The result of the comparison operation can be passed on to the digital layer.
- Digital layer: provides binary storage (long-term), logic operations and arithmetic operations (implemented using bit-serial approach). It also provides asynchronous propagation operations, global logic, pixel address extraction and high-speed I/O. The digital layer produces a 'flag' signal that selectively disables processors in the analogue layer.

As the through silicon vias still occupy considerable silicon area, it is beneficial to partition the cell across layers so that wiring requirements between the layers are minimised. In our design, the sensor layer communicates with the processor layer through a single wire, carrying out current value corresponding to the pixel brightness. The analogue layer receives input from the sensor layer, and communicates with the digital layer through a two-via interface (one input, one output). Control signals for all the layers are distributed on the respective layers, and the overall coordination of activity and communication is ensured through the synchronised control of all layers.

III. Circuit Design

A. Photosensor Layer

The first layer of the device is composed of an array of 32 by 32 current mode pixels. All pixels in the imaging array contain 3 transistors and a photo sensitive element within a pitch of 30μm by 30μm. In this imaging array, two different photo sensitive elements are implemented for comparative reasons: a photodiode and a photogate (sections of the array contain different pixel designs)

The optically transparent substrates of the SOI technology make vertical alignment of multiple dies economically feasible. The relatively small thickness of the substrate has been beneficial in the 3D integration process, alleviating some of the power concerns. On the same account, the thin substrate does not provide an optimum medium for light absorption. The typical absorption length for visible light in silicon is many microns, far greater than the thickness of the silicon film in typical modern fully-depleted (FD) silicon-on-insulator or silicon-on-sapphire processes. Previous works have focused on achieving respectable levels of quantum efficiency by using lateral PIN photodiodes with large photosensitive intrinsic regions, while balancing the tradeoff between quantum efficiency and pixel area [17, 18].

A schematic of the pixel with the two different photo sensitive structures is presented in Figure 2. The photodiode is fabricated in an annular "doughnut" shape to avoid edge effects, which can potentially increase dark currents. Poly silicon is placed on top of the entire photodiode structure in order to minimize lattice defects between the P+/N+ doped silicon and SiO_2 interface, which can also lead to higher dark

currents. The cross section of the photogate is presented on the left side in Figure 2. The gate of the photo transistor is externally biased in order to create a channel between the source and drain terminal. The channel will form a vertical photodiode with the P+ substrate which is backside illuminated though the 600nm thick SiO_2. The photocurrent is integrated on the floating diffusion node capacitance of the phototransistor's drain node.

The reset transistor M1 controls the operation mode of the photo sensitive element. When the gate of the reset transistor is set high, the photodiode/photogate is reset to Vreset potential. Once the photodiode is reset, the gate terminal of the reset transistor is set low and the photocurrent is integrated for one frame period (e.g. 33msec). The gate terminals of all reset transistors in the imaging array are connected together and allow for global integration and reset of all pixels simultaneously. Transistor M2 operates in the linear mode and converts the integrated photovoltage into an output current. Keeping the M2 drain potential constant at ~400mV (using transistor M3 whose gate is globally biased) allows M2 to act as a linear transconductor. Hence, a current output, linearly proportional to the integrated photo charge, is passed to the processing units in the subsequent layers.

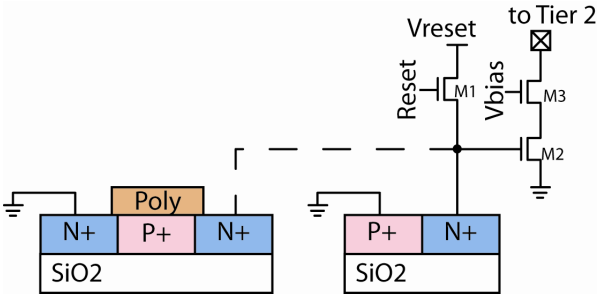

Figure 2. Two photopixels implemented in the first tier.

B. *Analogue Layer*

The second silicon layer contains the analogue processing elements (APE), shown in Figure 4. The APE array is dedicated to gray-level image processing operations. It consists of four analogue current-mode memory cells (one of them connected to four nearest neighbours), an input current cell, and a comparator circuit. All are connected to a common analogue bus through analogue switches. The photodetector current (from the top tier) is also selectively fed into the bus. The operation follows the general principles of instruction-level programmable analogue processing outlined in [19]. The memory cell uses a modified S^2I circuit. It operates with a 2.5uA reference current and 1.5V supply voltage. Body connected transistors have been used to increase the gate capacitance associated with the memory nodes and large-area transistors have been used to achieve good matching and minimize errors. The comparator provides its binary output signal to the tier below (digital processor). The flag signal (selectively disabling writing to the analogue registers) is produced by the digital layer.

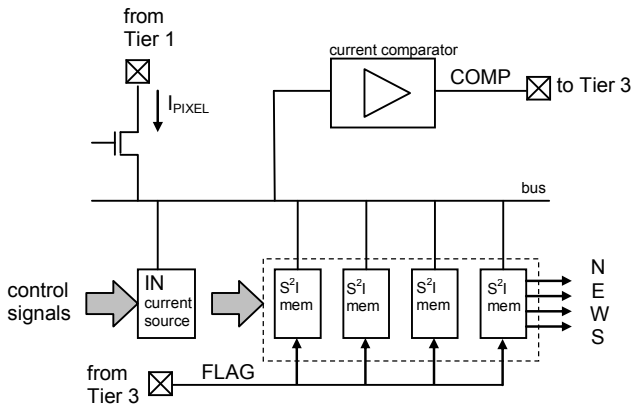

Figure 3. Analogue Processing Element (APE)

C. *Digital Layer*

The digital part of the design is allocated to the third tier. Structurally, it can be considered as separate sub-array, controlled by separate set of instruction lines, although in future, for optimization purposes, some instructions can be combined with analogue functions. This sub-array is dedicated to basic digital processing and simple global asynchronous operations. While analogue sub-array is effective in gray-scale arithmetic operations, digital array is used for storing and efficient processing of binary data, such as masks, markers and others.

Each digital processing element (DPE) is connected to its four neighbours for local data exchange. Every DPE contains 12 bits of local RAM, flag register and bus controller (BC). The general structure of the cell is illustrated in Figure 4. Operating as data-path all logic operations on local or neighbourhood data are performed on a bus or in BC respectively. Each cell facilitates the full set of Boolean operators.

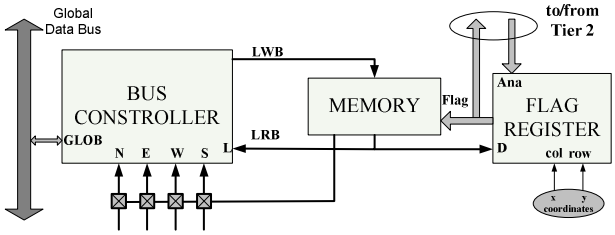

Figure 4. Digital Processing Element(DPE)

In order to introduce local autonomy features for both analogue and digital sub-arrays, each DPE incorporates a flag register. All memory operations in APE and DPE are controlled by a local flag indicator. The flag value can be either set to the value of Local Read Bus (LRB) from DPE (i.e. the value of local digital memory) or to the output of a current comparator in the APE. Additionally, the flag register is used to enable random access to PE's by their (x,y) coordinates (i.e. row and column select signals).

The bus controller acts as a multiplexer of internal, external or local neighbourhood data to the Local Write Bus (LWB). Functionally, the BC is an OR gate, thus enabling logic operations on global and neighbourhood data. Apart from that, BC facilitates asynchronous propagations of binary trigger waves across the entire array. The mask that defines whether the propagation is allowed or prohibited from certain direction is stored in the local memory. Hence, the network topology can be constrained locally. The produced binary trigger-waves can be efficiently used for such global operations as object reconstruction, hole filling, watershed transformation and many others [4].

In order to minimize the logic and the number of instruction lines used for addressing, the local memory is organised as 2D array of SRAM cells, addressed by two binary values. Previously, in order to comply with strict area constrains, we used an optimised dynamic latch as a basic memory unit [4]. However, in a SOI technology used for this design, dynamically stored charge becomes very vulnerable to noise from external logic and wiring, because of channel coupling and floating body. Therefore, in current implementation we use a more conventional SRAM cell, shown in Figure 5, to ensure the correct memory operation. This leads to a certain sacrifice in terms of memory size.

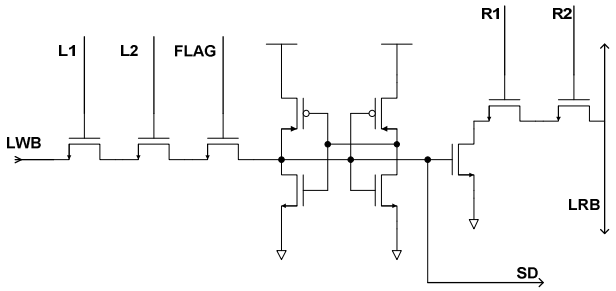

Figure 5. SRAM memory cell used in the digital layer.

IV. IMPLEMENTATION

All layers have been implemented with a 30 μm × 30 μm pixel pitch, The APE circuitry contains 56 transistors, and DPE circuitry consists of 157 mostly minimum-size transistors. The circuitry doesn't actually use all the available area and there is a scope for introducing additional functionality or increasing local memory. The performance of digital array is estimated according to executed binary operations. According to post-layout simulations, the correct operation is expected at frequencies up to 350 MHz. At this frequency 32×32 digital sub-array provides 179×10^9 binary operations per second. Furthermore, the overall binary performance can be significantly boosted by the gain from asynchronous processing, as DPE processing time equates to the signal propagation time through a combinatorial circuit of the BC. Analogue (gray-level) operations were designed to be executed at 10 MHz rate, and with the accuracy comparable with our previous design [3] (signal-dependent error below 1% and error matching better than 0.1%).

V. CONCLUSIONS

We have overviewed the architecture and circuitry of a 3D integrated pixel-parallel SIMD sensor/processor array chip. The vertical die stacking offers a very promising solution to the design of such devices, enabling optimized image sensing and processing, and the circuit expansion in the third dimension provides enhanced processor functionality and local memory capacity without introducing area penalty, enabling future high resolution processor-per-pixel devices. One can start to imagine how such architectures could be extended to 3D devices with tens and hundreds of tiers. The chip has been submitted for fabrication and the experimental results will be reported at a later date.

ACKNOWLEDGMENT

This research was sponsored in part by EPSRC grant no. EP/D029759/1. Fabrication was provided by MIT Lincoln Laboratory.

REFERENCES

[1] A. Moini, *Vision Chips*, Kluwer, Norwell, 2000
[2] P. Dudek, P.J.Hicks, "A general-purpose processor-per-pixel analogue SIMD vision chip", *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 52, no. 1, p. 13-20, 2005
[3] P. Dudek and S.J.Carey, "A General-Purpose 128x128 SIMD Processor Array with Integrated Image Sensor", *Electronics Letters*, Vol.42, No.12, pp.678-679, June 2006
[4] A. Lopich, and P. Dudek, "ASPA: Focal plane digital processor array with asynchronous processing capabilities", *ISCAS 2008,* p.1592-1595
[5] A. Lopich, and P. Dudek, "Asynchronous Cellular Logic Network as a Co-Processor for a General-Purpose Massively Parallel Array", *Int. Journal of Circuit Theory and Applications* (in press)
[6] T. Bernard et al., "A Programmable Artificial Retina". *IEEE J. Solid State Circuits*, Vol. 28, No. 7, pp. 789–797, July 1993
[7] A. Paasio et al, "A compact computational core for image processing," in *Proc. ECCTD 2001*, vol. 1, pp. 337–339, 2001.
[8] A. Rodríguez-Vázquez et al., "ACE16k: The Third Generation of Mixed-Signal SIMD-CNN ACE Chips Toward VSoCs". *IEEE Trans. on Circuits and Systems-I*, Vol. 51, No. 5, pp. 851-863, May 2004.
[9] L. O. Chua and T. Roska, *Cellular Neural Networks and Visual Computing.* Cambridge University Press, Cambridge UK, 2002
[10] L. Lindgren et al., "A Multi-resolution 100-GOPS 4-Gpixels/s Programmable Smart Vision Sensor for Multi-sense Imaging". *IEEE Journal of Solid-State Circ.*, Vol. 40, No.6, pp. 1350-1359, July 2004
[11] Chih-Chi Cheng et al., "iVisual: An Intelligent Visual Sensor SoC with 2790fps CMOS Image Sensor and 205GOPS/W Vision Processor". *ISSCC 2008*, pp.306-307, San Francisco CA, Feb. 2008
[12] J. Burns et al. "An SOI three-dimensional integrated circuit technology," *IEEE Int. SOI Conference*, October 2000, pp. 20–21.
[13] J.U. Knickerbocker et al., "3D silicon integration", *Proc. of ECTC 2008*, pp. 538-543, Orlando FL, May 2008
[14] V. Suntharalingam, et al., "Megapixel CMOS image sensor fabricated in three-dimensional integrated circuit technology," *ISSCC Dig. Tech. Papers*, pp. 356-357, Feb. 2005
[15] M. Koyanagy et al., 'Neuromorphic vision chip fabricated using three-dimensional integration technology', *IEEE International Solid-State Circuits Conference, ISSCC 2001,* Vol. 1, pp. 270–271, February 2001
[16] Zhengming Fu and E. Culurciello "A 3D Integrated Feature-Extracting Image Sensor", *ISCAS 2007*, pp.3964-3967, 27-30 May 2007
[17] A. Afzalian and D. Flandre, "Physical modeling and design of thin-film SOI lateral PIN photodiodes," *IEEE Trans. Electron Dev.*, vol. 52, no. 6, pp. 1116-1122, June 2005.
[18] C. Xu, C. Shen, W. Wu, and M. Chan, "Backside-illuminated lateral PIN photodiode for CMOS image sensor on SOS substrate," *IEEE Trans. Electron Dev.*, vol. 52, no. 6, pp. 1110-1115, June 2005.
[19] P. Dudek, P.J. Hicks, "A CMOS general-purpose sampled-data analog processing element", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol 47, no. 5, p. 467-473.